AFRL-HE-WP-TR-1998-0046

UNITED STATES AIR FORCE RESEARCH LABORATORY

BUILDING THE LEM²*R³ MODEL OF PILOT TRUST AND DYNAMIC WORKLOAD ALLOCATION

A Transition of Theory and Empirical Observations to Cockpit Demonstration

Peter G. Raeth

WRIGHT LABORATORY
WRIGHT-PATTERSON AFB 45433

John M. Reising

HUMAN EFFECTIVENESS DIRECTORATE
CREW SYSTEM INTERFACE DIVISION
WRIGHT-PATTERSON AFB OH 45433-7022

19980805 083

FEBRUARY 1998

FINAL REPORT FOR THE PERIOD JANUARY 1994 TO OCTOBER 1997

Human Effectiveness Directorate
Crew System Interface Division
Wright-Patterson AFB OH 45433-7022

DTIC QUALITY INSPECTED 1

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.

- Pages with background color or light colored printing.

- Pages with small type or poor printing; and or

- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☑ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

> National Technical Information Service
> 5285 Port Royal Road
> Springfield, Virginia 22161

Federal Government agencies registered with the Defense Technical Information Center should direct requests for copies of this report to:

> Defense Technical Information Center
> 8725 John J. Kingman Road, Suite 0944
> Ft. Belvoir, Virginia 22060-6218

DISCLAIMER

> This Technical Report is published as received and has
> not been edited by the Air Force Research Laboratory,
> Human Effectiveness Directorate.

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-1998-0046

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

HENDRICK W. RUCK, PhD
Chief, Crew System Interface Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1998	Final, January 1994 - October 1997 1997

4. TITLE AND SUBTITLE

Building the LeM^2*R^3 Model of Pilot Trust and Dynamic Workload Allocation

5. FUNDING NUMBERS

PR: 2403
TA: 04
WU: 86

6. AUTHOR(S)

* Peter G. Raeth
John M. Reising

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

* Wright Laboratory
Wright-Patterson AFB OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory (AFRL), Human Effectiveness Directorate
Crew System Interface Division
Air Force Materiel Command
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

AFRL-HE-WP-TR-1998-0046

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

For pilots to accept active decision aids during complex flight scenarios, it is essential that the automation work is in synergy with aircrew. To accomplish this, the automation must go well beyond menu and macro selections, where the pilot must explicitly tell the automation what to do and when to do it. It must also transcend "mother may I" approaches, where the automation asks for permission to proceed. To these traditional barriers, the automation needs a sense of how the pilot will react in a given situation and, based on that reaction, how much of the workload could be allocated to the automation at any given time. For this purpose, the authors reviewed the literature on human factors and dynamic function allocation. This literature provided a wealth of information on this topic. Based on the current state of the art in this topic area, the authors developed and tested a dynamic model of pilot trust and workload allocation. This "full degrees of freedom" model transitions human factors theory, as it exists today, into an engineering application. The resulting model can be combined with other information obtained from static and continuous processes to divide the workload and minimize cognitive overload.

14. SUBJECT TERMS human factors, function allocation, cockpit design, decision aids, pilot associate, artificial intelligence, uninhabited combat aerial vehicle, UCAV, UAV, operator aids, pilot aids, function estimation, parameter estimation, simulation modeling, human-computer interface, human-computer interaction

15. NUMBER OF PAGES
99

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

Appendices:

LIST OF FIGURES

LIST OF FIGURES (Continued)

ACKNOWLEDGEMENTS

THIS PAGE INTENTIONALLY LEFT BLANK

1.0 INTRODUCTION

For pilots to accept active decision aids during complex flight
scenarios, it is essential that the automation work in synergy with
the aircrew. To accomplish this, the automation must go well beyond
menu and macro selections, where the pilot explicitly tells the
automation what to do and when to do it. It must also transcend
"mother may I" approaches, where the automation asks for permission
to proceed. To breach these traditional barriers, the automation
needs a sense of how the pilot will react in a given situation and,
based on that reaction, how much of the workload could be allocated
to the automation at any given time. To advance the applied aspects
of the research in this area, the authors transitioned results from
three major research efforts. We implemented a dynamic model of
pilot trust and workload allocation that correlates with the human
data collected during those efforts. This "full degrees of freedom"
engineering model is the beginning of building a continuous
adaptive process that divides cockpit workload and minimizes
cognitive overload.

1.1 BACKGROUND

Aerospace operational requirements derive from complex scenarios
that tax the cognitive powers of the best people. Exacerbating the
situation is the trend to decrease the number of pilots in the
cockpit while expecting equal or improved mission performance.

Thus, automation has been expanded to enable success in spite of potentially intractable scenarios. Figure 1 illustrates this idea. The "electronic crewmember" observes the situation, integrates relevant information, and takes on workload as appropriate. Thus, the pilot is free to act as a manager focused on mission goals instead of being overburdened with details. In this way, performance requirements can be met in spite of cognitive barriers.

For the automation to be a successful part of the system given these conditions, the automation must be dynamic, and work in synergy with the pilot. It is more than ego that demands that pilots always be in charge. Pilots are the reason for and the primary part of the system. It is pilots who have the ultimate responsibility for accomplishing the mission. Aircraft are the pilots' tools. Putting pilots out of the decision-action loop surrenders control and defeats the system's purpose. Still, some automation independence is needed given scenario complexity and pilot cognitive limits. One factor to consider, in both static and dynamic workload allocation, is a prediction of pilot trust and its impact on workload sharing.
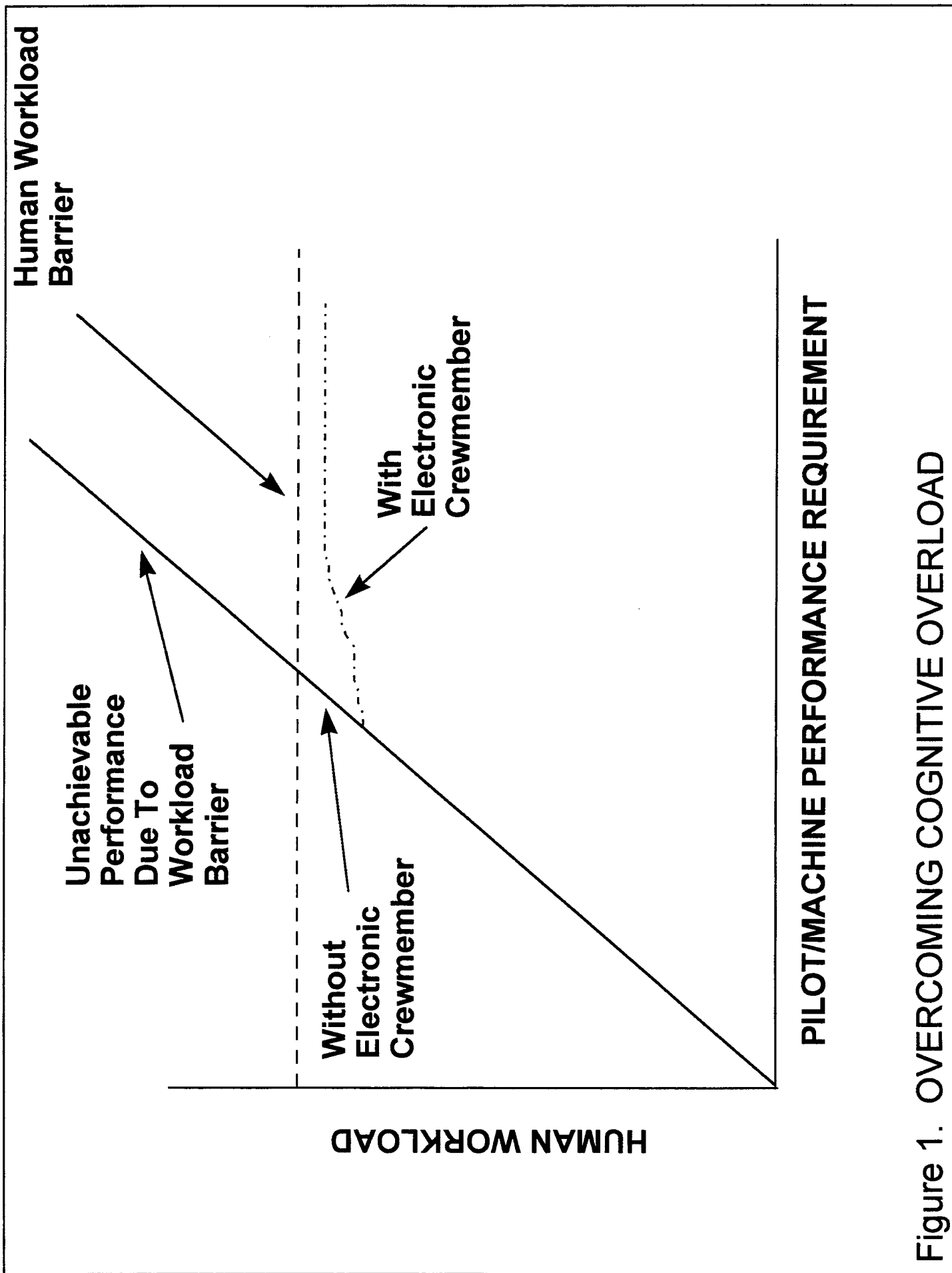
Figure 1. OVERCOMING COGNITIVE OVERLOAD

1.2 PREVIOUS RESEARCH

The Air Force Pilot's Associate program developed two approaches to workload allocation. One approach employed static workload allocation methods whereby the crew defined, pre-mission, what workload the automation should undertake during specific mission phases [Lilley, 1995]. [1] The other approach applied a real-time estimate of pilot workload [Riley, 1997]. The pilot could deny any workload allocation proposal made by the computer. While the static approach to allocation can provide a good baseline, a dynamic adjunct is needed that can adjust the workload depending on unforeseen circumstances. A real-time estimate of pilot workload is a valuable approach to accomplishing this. Adding to the field are Pope and Bogart who developed a continuous workload adjustment capability based on measurements of brainwaves. Morrison's team contributed to this area by developing dynamic allocation design guidelines. They also developed a means of varying task allocation based on workload projections.

Yet another thrust is represented by this report on a method of adjusting workload based on pilot trust, historical system performance, and scenario conditions as they evolve during a mission. We transitioned the seminal work of Lee and Moray [1992 and 1994] and tested the resulting dynamic engineering model of

---

[1] In all references to "function allocation" or "workload allocation", the pilot allocates to the automation.

4

pilot trust and workload allocation. [2] This model can be combined with other information obtained from static and continuous processes to dynamically allocate tasks between the pilot and the automation. The result can be installed in a cockpit evaluation system for pilot-in-the-loop part-task testing. The first attempt to build, test, and apply this model was reported by Raeth, Noyes, and Montecalvo. Later efforts are reported by Raeth and Reising.

This new model adds to Lee and Moray's model in the following ways:

a) the dimension of danger as experienced during combat was added to their model (an insight derived from Riley, 1994)

b) their fault (system error) parameter was expanded from four discrete values to continuous values over the entire range 0-100%

c) their concept of history (time at t-1) was expanded beyond the just-past sample to a weighted moving average taken over the last n samples. (The weighting scheme varies according to the type of pilot, as explained later.) [3]

---

[2] What is reported here is a view of the human factors literature as seen from the computer engineering world.

[3] It is important to note that Lee and Moray had excellent reasons for limiting their model to t-1. It was not an arbitrary choice. Their model is derived directly from empirical data. They used time series analysis software to develop a model that best fit their data. This approach tries to fit the model using only t, then t and t-1, then t and t-1 and t-2, and so on. The software provides information so that it is possible to detect how many terms make a difference. In their case, going from t-1 to include t-2 or t-3 did not make the model fit the data any better. t-2 and beyond were just random noise down in the residuals. Hence, they concluded that, empirically, humans, in their experiments' situation, do not use more than "one back" [Moray, 1997].

We added the extra time history because we felt this was necessary to predict the range of responses resulting from the dimensions of danger and various pilot types. This addition to the model was based on our understanding of ideas found in the human factors literature (see the appendix). We took this track because the benefits of dynamic workload allocation need to be demonstrated in

a military cockpit emulation system. Such a demonstration is necessary to encourage continued support for the research. An added benefit of the demonstration would be the feedback from the military pilots who participate. With continued support, additional empirical data can be gathered and the model improved. To create a combat demonstration, it was necessary to look to the literature as a guide to implementing the results from empirical experiments performed in benign process observation and control environments.

2.0 AN EXAMPLE COMBAT SCENARIO INVOLVING DECISION AID FAULTS

As we go into detail on how dynamic function allocation might be implemented - it will be easier to understand if we apply it to a combat scenario. The key purpose of the scenario is to illustrate that decision aids (automated aids for the pilot) are not perfect, they make mistakes (faults). The mistakes a decision aid makes have an impact on pilot trust and, subsequently, on the workload "he" is willing to allocate to the automation, either statically pre-mission or dynamically as the mission proceeds.

2.1 DECISION AIDS AND FAULTS

There are three basic types of decision aid faults. The first type results from the fact that decision aids are primarily heuristic, not algorithmic. Thus, their goal is to give acceptable, although not necessarily optimal, advice. As a result, they sometimes give, for instance, false alarms or poor route plans. The second type of decision aid fault is caused by a hardware failure. An example is a circuit breaker activation that prevents certain subsystems from operating correctly, if at all. This fault type could disable sensor, radar, communication, processor, display, and other critical subsystems. Such faults would cause the decision aid to malfunction or deliver incorrect information. The third fault type results from software coding errors, database entry errors, data communication errors, and specification errors. These cause malfunctions even though the hardware is operating correctly.

7

Decision aid faults are an excellent means for illustrating situationally-based trust and the subsequent impact on the percentage of workload the pilot is likely to allocate to the automation. For our purposes, all decision aid faults are assumed to be equally significant. This may not be the case in an actual situation. The assumption is made merely for demonstration purposes.

2.2 AIR-TO-GROUND SCENARIO DESCRIPTION

Let us consider a typical single-seat, ground-attack mission against enemy high-value targets. The mission actually starts about three hours before take-off time with mission briefing and planning. At this stage, crucial routing decisions will be made depending on topography over the route, other friendly forces that have been committed, and intelligence on the position of enemy formations. It is now that many of the preconceptions of the mission that will later color decisions are formed, and yet it is at this stage that automation has been incorporated to a very large extent with the introduction of computer-driven flight planning, threat warning, and threat response systems. It is important that the synergy between operator and machine begin at this point.

The first phase of flight consists of loading the aircraft with pre-flight data prepared during mission planning. This is followed by take-off (mission phase **T.O.** in Figure 2) and then a transit toward the target area, usually at high to medium altitude, in

8

preparation for descent to low altitude, prior to entering enemy territory. This phase may include several rendezvous requirements for force gathering, air-to-air refueling, etc. Generally, however, it is a time of relatively low workload and danger, but very high anticipatory stress.

In order to reach the target, the aircraft must cross enemy territory. This typically necessitates crossing the region where the rival land forces are in conflict. This is generally known as the Forward Edge of the Battle Area (**FEBA**). With descent to low altitude and penetration of the FEBA comes the first peak of workload and a radical rise in danger. Upon entering the FEBA, the aircraft may be anywhere between 50 and 500 feet above the ground, probably at night, and in less than optimal weather conditions. The pilot is relying on a number of complementary systems working in the visible, infra-red, and radar frequencies. All of these systems have unique properties in terms of the information provided. But, none of these systems are capable of providing sufficient information in isolation. In addition to the basic flying task, the pilot is passing over territory with a densely packed assortment of threats arrayed in a confusing and constantly changing tactical situation.

Figure 2.
Fighter's Air-to-Ground Scenario

THERE ARE SEVEN MISSION PHASES:

Event	Time Units	Minutes Into Mission	Description
T.O.	000	0	Take off
FEBA	125	25	Cross FEBA into enemy territory
ING	150	30	Fly toward target area
TGT	375	75	Enter target area
TGT	525	105	Depart target area
EGR	575	150	Fly away from target area
FEBA	775	155	Cross FEBA into friendly territory
LND	900	180	Land

THERE ARE FOUR SYSTEM EVENTS:

Event	Time Units	Minutes Into Mission	Description
F 1	400	80	First decision aid fault occurs / First system performance loss
R 1	425	85	Full recovery
F 2	525	105	Second decision aid fault occurs / Second system performance loss
R 2	550	110	Full recovery

Danger

Danger (% of Maximum)

Time

In order to make adequate sense of the environment, it is necessary to compare data being received from on-board sensors with data link information being received from control centers, satellites, and other aircraft. This must then be compared with on-board data bases and mission planning requirements in order to identify any possible conflicts and to develop a coherent picture of the world and its threats, as perceived by the aircraft. This, in turn, must be collated into a coherent strategy for the successful completion of the mission in terms of an optimum route and countermeasures deployment. The data from this myriad of sources is fused into a minimum data set necessary for the pilot to perform the task at hand. By pilot request, or prior agreement, the automation will perform certain tasks directly. This serves to minimize pilot workload and permits the pilot to manage the mission, rather than be burdened with low-order details.

Having successfully negotiated the FEBA, the next phase is a low-level ingress to the target site (ING). It is during this phase that the workload varies the most and that danger continues to rise. In the ideal scenario, the pilot avoids all enemy defense activity, and the challenge comes from physically flying the aircraft at dangerous altitudes in appalling weather at night (if no automatic terrain following system is employed). In the real world, the pilot will likely be presented with a series of unexpected situations arising from enemy action (threats and air-defense fighter activity) or possibly his own system malfunctions

that will require re-evaluation of the flight plan and mission capability (system events Failure_1 (**F1**), Recovery_1 (**R1**), Failure_2 (**F2**), and Recovery_2 (**R2**) in Figure 2). In our case, these events occur while the pilot is in the target area.

The mission workload and danger rise to a crescendo over the target area (**TGT**), where aircraft system activity for weapon deployment, a concentration of enemy defense activity, coordinated attack timing considerations, and precision flying requirements for weapon targeting all compound the operator's task.

Egress from the target site (**EGR**), the flight through enemy territory, crossing the FEBA and, ultimately, recovery to base (**LND**), present essentially the same set of problems and fluctuating workloads to the pilot as do the mission phases prior to target attack. However, they are compounded by fatigue and possible battle damage sustained by the aircraft during the sortie.

2.3 SCENARIO SUMMARY

As illustrated in Figure 2, this scenario consumes 900 time units. The flying part of the mission takes three hours, making each time unit occur at 0.2 minute intervals, or every twelve seconds. A new observation of the situation, aircraft systems, and pilot is produced each time unit. As described in the scenario above, the mission goes through seven phases during which two system faults occur. One fault happens just as the pilot enters the target area.

12

The other takes place as the pilot departs the target area. The mission phases and system faults are summarized in the table below.

THERE ARE SEVEN MISSION PHASES:

Event	Time Units	Minutes Into Mission	Description
T.O.	000	0	Take off
FEBA	125	25	Cross FEBA into enemy territory
ING	150	30	Fly toward target area
TGT	375	75	Enter target area
TGT	525	105	Depart target area
EGR	575	150	Fly away from target area
FEBA	775	155	Cross FEBA into friendly territory
LND	900	180	Land

THERE ARE FOUR SYSTEM EVENTS:

Event	Time Units	Minutes Into Mission	Description
F 1	400	80	First decision aid fault occurs First system performance loss
R 1	425	85	Full recovery
F 2	525	105	Second decision aid fault occurs Second system performance loss
R 2	550	110	Full recovery

Now that the scenario mission phases and system events have been discussed, we will illustrate a means to predict pilot trust and workload allocation. This would permit the workload to be dynamically adjusted without exceeding the pilot's comfort level with the automation. In this way, the best synergy between pilot and automation might be achieved. The technique proposed here is based both on experimental observations and some basic ideas in human/computer trust discussed in the appendix.

13

(This Page Left Blank)

3.0 THE ORIGINAL TRUST MODEL

Lee and Moray [1992] published a model based on their experiments where subjects operated a process control simulation. The subjects rated their trust in the system during various modes of operation and the resulting data was analyzed. The model they derived from their analysis is stated below. They derived their model by first collecting empirical data and then applying an autoregressive moving average vector form of time series analysis.

$$\text{trust}_t = 0.570\text{trust}_{t-1} - 0.740\text{fault}_t + 0.740(0.400)\text{fault}_{t-1} + 0.062\text{performance}_t - 0.062(0.210)\text{performance}_{t-1}$$

Where: trust ranges from 1 to 10 and is the level of trust the operator has in the system, level 1 is the lowest

performance ranges from 0 to 100 and is the percentage of the best system performance relative to mission objectives; "system" refers to the combination of operator and supporting automation;

fault represents the fractional variation of a control system variable vs. that set by the operator (i.e. the pilot sets 14 and the system delivers {(1.0 +/- fault) * setting}; originally, this variable was used to represent four discrete coded percentages: 1 (15%), 2 (20%), 3 (30%), and 4 (35%)

t refers to the current value of the variable

t-1 refers to the most recent historical value of the variable

At issue with this model, if one were to use it to predict a pilot's reaction during a complex flight scenario, is the range of fault and performance values tested by Lee and Moray during their original experiments. To work well in a cockpit application, the

15

model needs to be extended to account for the full range of these two parameters. Essentially, one needs each parameter to vary from its minimum to its maximum on a continuous scale.

Further, when one reads the work of Riley, one realizes that the danger of the situation has an impact on the pilot's reaction. In his 1989 paper, he proposed the theoretical impact of danger. In his 1994 paper, he confirmed it experimentally.

Finally, Riley [1997] has observed that his simulation and the process control simulation from which the original trust model was derived were conducted at a much slower pace and the decisions were more discrete than would occur in the second-by-second, continuous evolution of a military aviation scenario.

4.0 EXTENSIONS TO THE ORIGINAL TRUST MODEL

To extend Lee & Moray's original trust model [1992], the authors applied human factors concepts and interpolation to derive fully continuous parameters from the discrete ones noted in the previous section. At the heart of this new model lies the original model. This new model shares characteristics with Riley's 1994 work and adds to Lee and Moray's model in the following ways:

a) the concept of history (time at t-1) was expanded beyond the just-past sample to a moving average taken over the last n samples.

b) their fault parameter was expanded from four discrete values to continuous values over the entire range of 0 to 100%

c) the dimension of danger experienced in a combat situation was added to the model

4.1 PARAMETER HISTORIES

The historical parameter values were originally those of the most recent sample (for instance fault at time t-1). However, the authors felt that the theory of human trust supported the notion that the reactions of different pilots would vary according to their experience and personality, especially in highly dangerous situations. Thus, some variable length of parameter history will be remembered by the pilot. The pilot's knowledge of the system's past operation and the use of that knowledge will impact "his" reaction to the system's current state. For instance, a certain amount of history will be remembered and the importance of recent memory will vary. To model the impact of memory length and to provide for the

17

importance of memory relative to the current time, historical values are taken to be a weighted moving average of n-number of past samples, instead of just the most recent sample at time t-1. The weighting scheme can also be varied according to the pilot being modeled. We cite three major schemes:

a) Constant Weight (Unweighted) - Every bit of information is equally important, no matter how old the information is. This scheme would reflect a pilot of infinite experience relative to the next two schemes. Since this is not very realistic, this scheme was not employed during testing and is mentioned only for the sake of completeness, in case it might prove useful in the future.

b) Linear Declining Weight - More recent information has greater importance and that importance declines gradually over time. This scheme reflects an **experienced** pilot who takes a balanced view of the system's past operation when reacting to the present. Such a pilot will not react too strongly or suddenly to present system operation. The less the slope of the line, the more experience the pilot has. That experienced pilots will tend to use the automation more than inexperienced pilots, even in case of failure, is partially supported by Riley's 1994 work.

c) Sigmoid Declining Weight - More recent information has exponentially more importance than older information. This scheme reflects an **inexperienced** pilot who reacts very suddenly and strongly to the system's present operation. This pilot does not take a balanced view of the system, taking little account of the overall system operation when reacting to the present. This pilot will tend to forget that use of the automation is necessary in order to successfully handle the excessive workload. The less the slope of the curve, the more experience the pilot has. Lee [1997] has observed that the original model's use of just the past sample at t-1 was established based on the fact that the samples at t-1-x had such little effect. This is similar to a sigmoid that decays extremely rapidly such that the values of all samples earlier than t-1 are essentially zero. In fact, the subjects used in the original experiments were inexperienced.

For our purposes, the pilot is assumed to have perfect memory. What varies is how much information is remembered (what is remembered is remembered perfectly) and how fast the importance of that information declines. By varying the length of the moving average, the type of weighting, and the maximum/minimum values of the weighting scale; various pilots can be accommodated. The assumption of perfect memory can be removed by perturbing the values stored in memory. This perturbation should probably be non-uniform, perhaps linearly decreasing, to account for recency effects.

This weighted moving average scheme for ranking importance depends only on when the information occurs, not on what type of information it is. Riley [1997] has observed that "Another possibility is that different information varies in importance to the pilot. More important information will be retained longer than less important information, and the pilot will respond more quickly to it. For example, if the pilot can do the automated task manually, so the automation's primary benefit is as a workload reducer, failure of the automation will be more important than recovery. Hence, the pilot will shut the automation off quickly if it fails, and turn it back on more slowly when it recovers." Riley's empirical evidence supports this idea.

Appendix A derives the equations we used to generate the weights.

4.2 EXTENDED ENCODING OF FAULT SIZE

Lee [1995] reported that the fault sizes used in his experiments were coded as four discrete fault sizes as represented in the following table. (Remember that the fault size is the difference between what the pilot wants the decision aid to do and what it actually does. In other words, fault size is the error the decision aid makes.)

Fault Size (as a % of maximum)	Code
15	1
20	2
30	3
35	4

Later, after this period of research was completed, we discovered

that the coded fault sizes were the mean of a random distribution. So, the experiments did not just use one fault size but a continuous range with the coded fault size as the mean. This new understanding can be incorporated into future versions of the model. According to Moray [1997], "... what we used as faults was a random variation about the value demanded, plus a constant. ... our faults were drawn from a continuous distribution. But, for the time series modeling we coded this as a discrete variable because the actual values the operators saw was a continuously varying function of time, and no one value, except the mean, represented the fault."

The codes are the values actually used in the equation. In order to establish a coding scheme that delivers continuous values from 0 to 100%, two additional codes were added by the authors.

Fault Size (as a % of maximum)	Code
0	0
100	5

The original and extended coding tables are compared in Figure 3. Linear interpolation is used to obtain the correct analog (continuous value) fault size code given a particular percentage of maximum fault size. The scheme used for linear interpolation is the same as that for obtaining history weights from a linear scale. This interpolation scheme can be easily expanded as more experimental data are collected.

Figure 3b. Extended Continuous Fault Size Codes

Figure 3a. Original Discrete Fault Size Codes

22

We took the step of converting the discrete codes to continuous ones at the risk of confusing interval data with ordinal data. The conversion was necessary from an engineering point of view because of the continuous nature of aerospace applications. More than one commentor has raised issue with Code 5 being 100%. Lee [1997] recommended that it be lowered to 45% since that value is a more likely extrapolation of his original data. The value was left as is for these first experiments since we felt we had to be able to accommodate a complete automation failure. Additional work is needed to determine how best to do that while maintaining a better link with Lee and Moray's original experiments, and those of Riley. The authors grant that this extension to the original table is not supported by experimental data. Still, the overall model results seem to track well with what we know about human behavior. Most troubling about the extension is the steep slope between codes 4 and 5. 65% of the fault percentages fall in that range. However, the original model was based on codes no larger than 4 and trial runs with that model showed that it was best to not let the maximum code go too far beyond 5. This is easier to accept if one assumes that Code 4 (35% system failure) is the threshold beyond which the system is essentially worthless. From the authors' experience in numerous experiments with novel vehicle-pilot interfaces, this is a safe assumption for aerospace scenarios.

The reader should understand that the current model represents a preliminary step, not a final postulation. Our goal is to implement the model in a cockpit evaluation system for pilot-in-the-loop part-task testing. Another goal is to provide a platform for continuous improvement and demonstration. The current implementation has proven to be a good start in that direction. Additional experiments and an analysis of the original data, using percentages instead of codes, would serve to "tighten up" the model's use of fault size coding.

4.3 THE DANGER PARAMETER

When one reads Riley [1994] and considers his background, one is led to realize that increasing danger has a negative impact on the pilot's willingness to use automation. "... risk may influence automation use decisions, with less reliance on automation when the possible consequence of errors is more severe and the automation has proven itself unreliable." (Riley, 1994, p116)

According to Reising, PERCENT DANGER = 100.0 x HAZARD x RISK. Where HAZARD is the probability that a set of circumstances could cause injury or death. RISK is the probability of occurrence of a hazardous event. Both values range from 0.0 through 1.0. For example:

if HAZARD = 0.90 and RISK = 0.50,

then DANGER = 100.00 x 0.90 x 0.50 = 45.00.

The authors' conjecture (partially supported by Riley's 1994 work) is that the maximum height to which trust will recover is inversely related to the size of the danger. Many equations could be proposed to give that effect to the model. The authors have chosen the following equation:

$RecoveryFactor=$

$$100.0 - \frac{\frac{FaultHistory+(100.0-PerformanceHistory)+DangerHistory}{3.0}}{10.0}$$

The parameter "RecoveryFactor" ranges between 0.0 to 10.0 and is mapped to the range 5.0 to 10.0 since that is taken as the trust recovery range. This mapped value is taken as the maximum height to which trust will recover. All trust values reported by the model are mapped to the range 1 through "MaxRecoveryHeight". The model's mathematics, however, are driven by the unmapped trust values. This ties the results more directly to the model's original formulation. An additional value of this equation is that trust does not recover as quickly nor as fully once it is lost (another insight from Riley [1989]). A final point is that the denominator 3.0 in the numerator or RecoveryFactor implies an equal weighting of the averaged parameters. We stayed with this weighting scheme since no justification could be found in the literature for doing otherwise. This initial setting may change as more empirical experiments are conducted and human factors theory advances.

5.0 STATEMENT OF THE FINAL TRUST MODEL

All the changes described above do not affect the previous statement of the model. However, a new understanding of the parameters is needed.

$$trust_t = 0.570 trust_{t-1} - 0.740 fault_t + 0.740(0.400) fault_{t-1} + 0.062 performance_t - 0.062(0.210) performance_{t-1}$$

Where:

trust is mapped to the maximum recovery height

performance is still 0-100%, as before

fault now takes on continuous values between 0 and 5

t still refers to the current time

t-1 now refers to the n-length weighted moving average of the parameter

(This Page Left Blank)

6.0   TESTING THE FINAL TRUST MODEL

At this point, it would be useful to illustrate the trust model's operation during the scenario described in Section 2.0.

6.1   DUAL-FAILURE SCENARIO TEST RESULTS

The results of running this scenario through the trust prediction equation are shown in Figures 4 and 5 for the experienced pilot (linear weighted history) and the inexperienced pilot (sigmoid weighted history). A summary comparison is given in Figure 6.

The simulation predicted that both pilots would lose trust more so during the second failure than during the first. It also predicted that the inexperienced pilot would suffer a more extreme loss of trust during each failure (down to 45%, then 42%) than would the experienced pilot (down to 46%, then 45%). In the short run, the experienced pilot was predicted to recover trust to a higher degree (87%) than would the inexperienced pilot (84%). In the long run, once the failures were cleared, their recovery rates were predicted to be nearly the same. The predicted differences in trust loss between the experienced and inexperienced pilot may be explained by the fact that the inexperienced pilot reacts far more strongly to the present situation. Later work with this dual-fault scenario needs to include Monte Carlo tests to determine if the differences predicted for the inexperienced and experienced pilots are statistically significant.

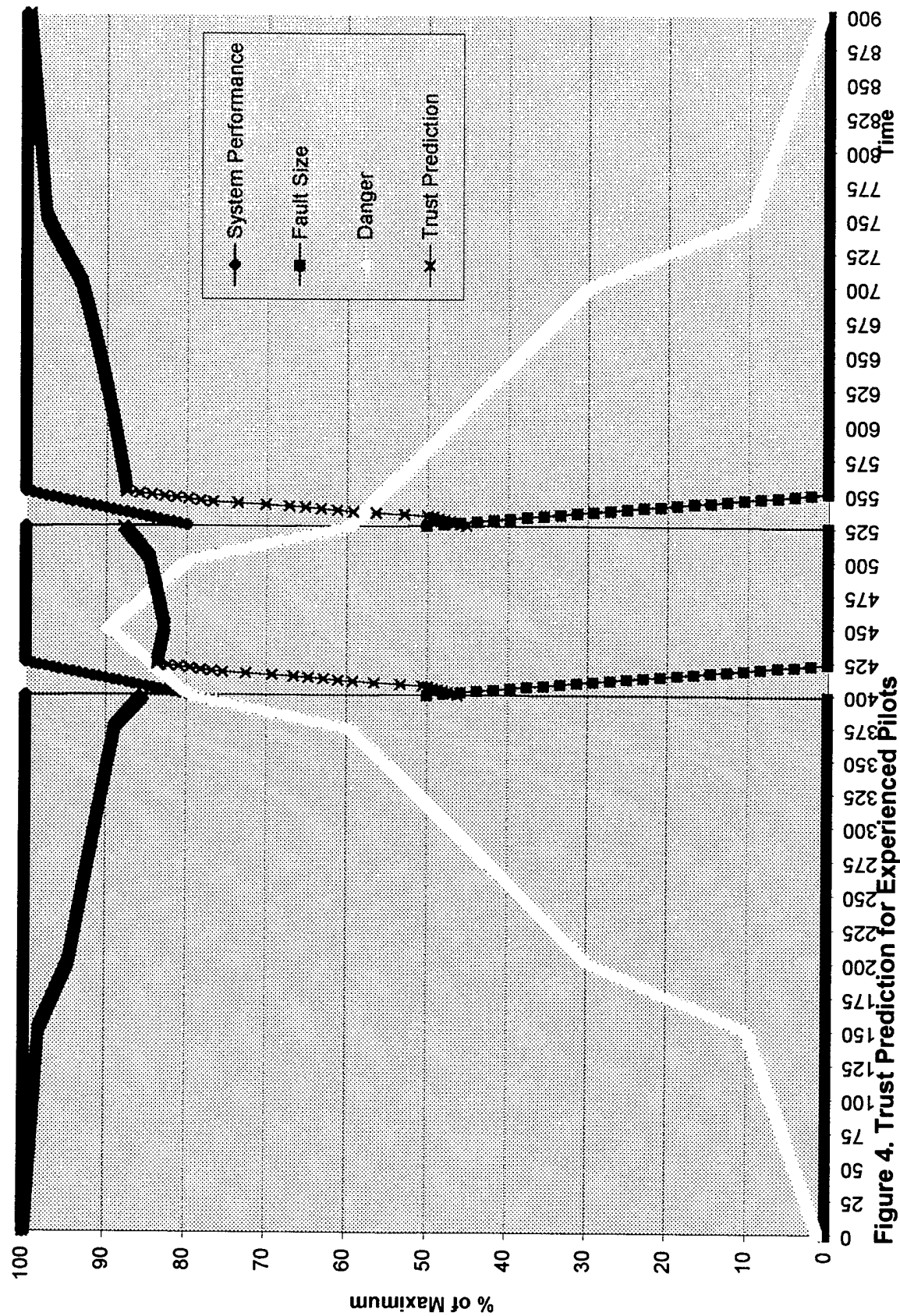Figure 4. Trust Prediction for Experienced Pilots
(Dual-Failure Scenario)

Figure 5. Trust Prediction for Inexperienced Pilots (Dual-Failure Scenario)

Figure 6. Trust Comparison: Experienced vs. Inexperienced Pilots
(Dual-Failure Scenario)

32

6.2 IMPACT OF MULTIPLE FAILURES

An additional test of the trust equation was performed using seven faults. Figures 7 and 8 show the predictions for the experienced pilot (linear weighted history) and the inexperienced pilot (sigmoid weighted history) respectively. Figure 9 compares them.

Note that the multiple-failure scenario yields predictions similar to those of the dual-failure scenario. With each failure, trust is lost to a greater degree by both pilots. This is due to the increasing danger and to the impact of historical system performance and faults. The inexperienced pilot is predicted to suffer a more extreme trust loss (at first, down to 42% and finally to 14%) compared to the experienced pilot (at first, down to 45% and finally to 19%). As before, in the short run, the experienced pilot is predicted to recover trust to a higher level (69%) than the inexperienced pilot (64%), once the failures cease. The short-term effect is due to the fact that the experienced pilot did not lose as much trust. (In the dual-failure case the experienced and inexperienced pilots both were predicted to recover 90% trust at nearly the same time, 627 and 655 respectively, since there were only two errors.) However, in the long run, the inexperienced pilot is predicted to recover trust much faster than does the experienced pilot (note the cross-over point at Time = 625). The experienced pilot reaches 90% trust at time index 816, whereas the inexperienced pilot reaches 90% at 731. The long-term effect is expected since the inexperienced pilot reacts more strongly to the

33

present situation than does the experienced pilot.

Riley's [1994] experiment #4 showed that trust is not affected by danger until after the second failure. However, our scenario assumes a much higher level of danger than did Riley's. Additionally, the pilot is assumed to know that much greater danger exists at different times in the scenario. We also assume that, while more dangerous times in the mission can be predicted in some general sense, the pilot does not know precisely when sudden great danger will occur. Thus, at least for this first engineering implementation, we are satisfied that the results extrapolate well from the empirical data.

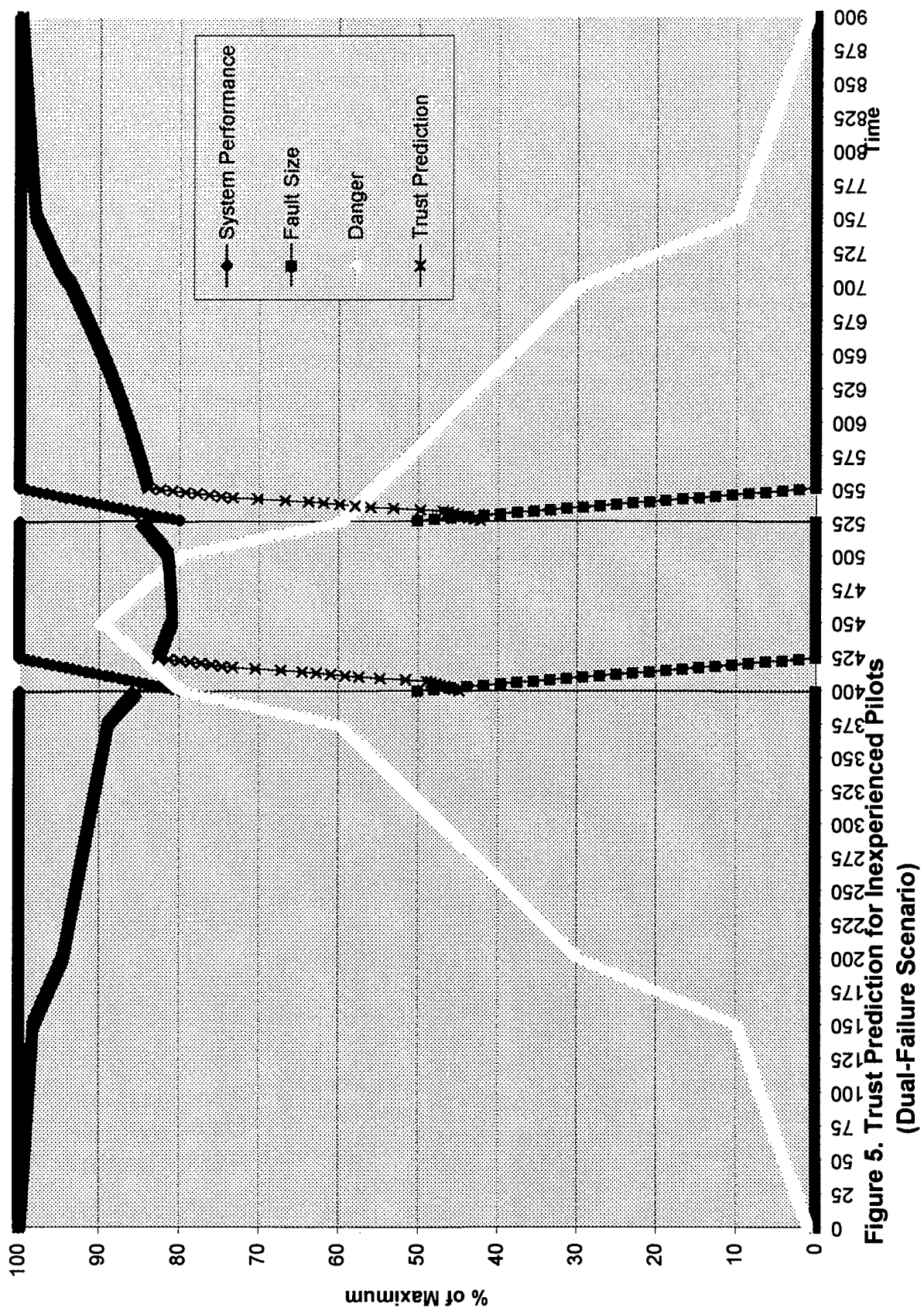Figure 7. Trust Prediction for Experienced Pilots (Multiple-Failure Scenario)
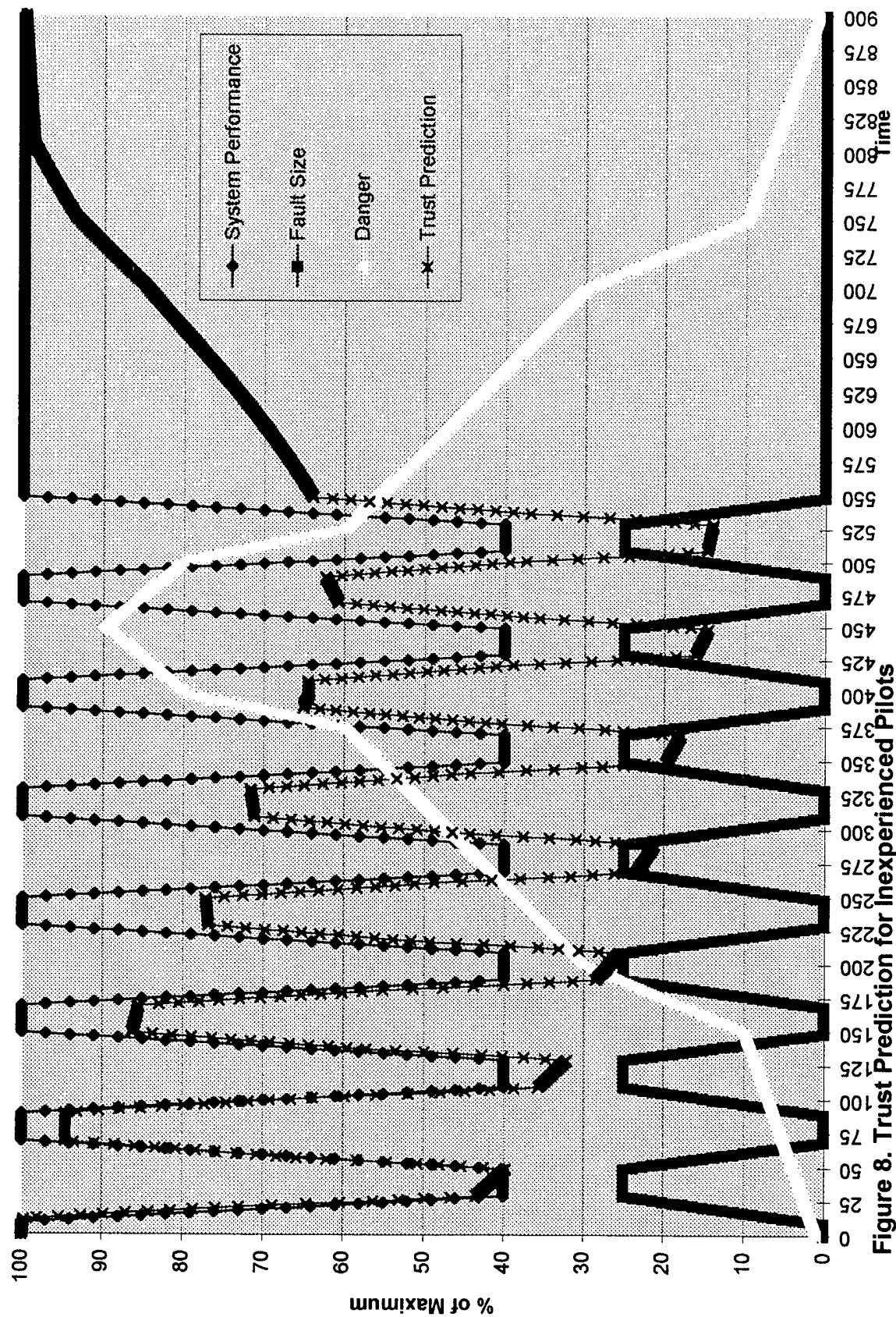
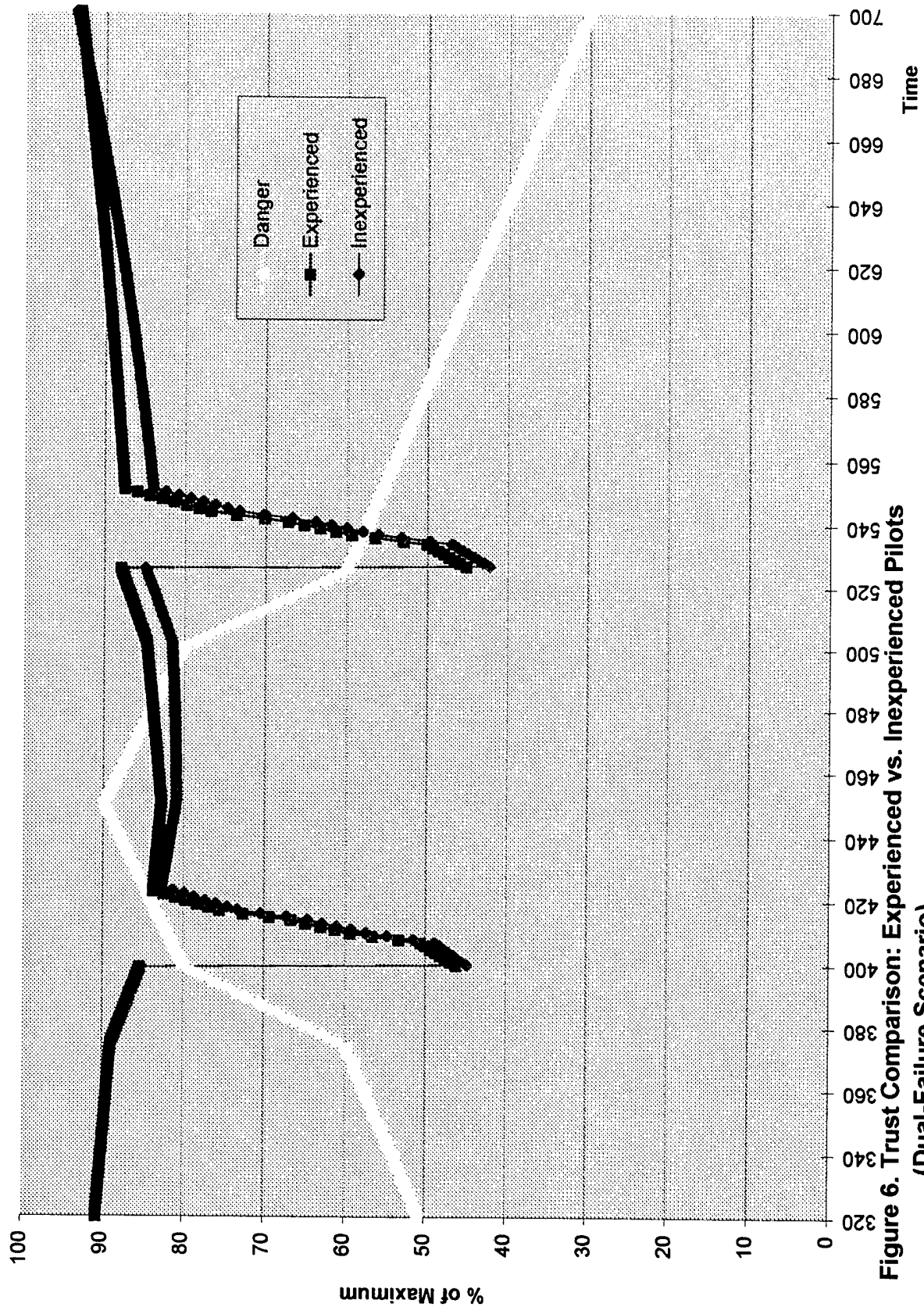Figure 8. Trust Prediction for Inexperienced Pilots (Multiple-Failure Scenario)

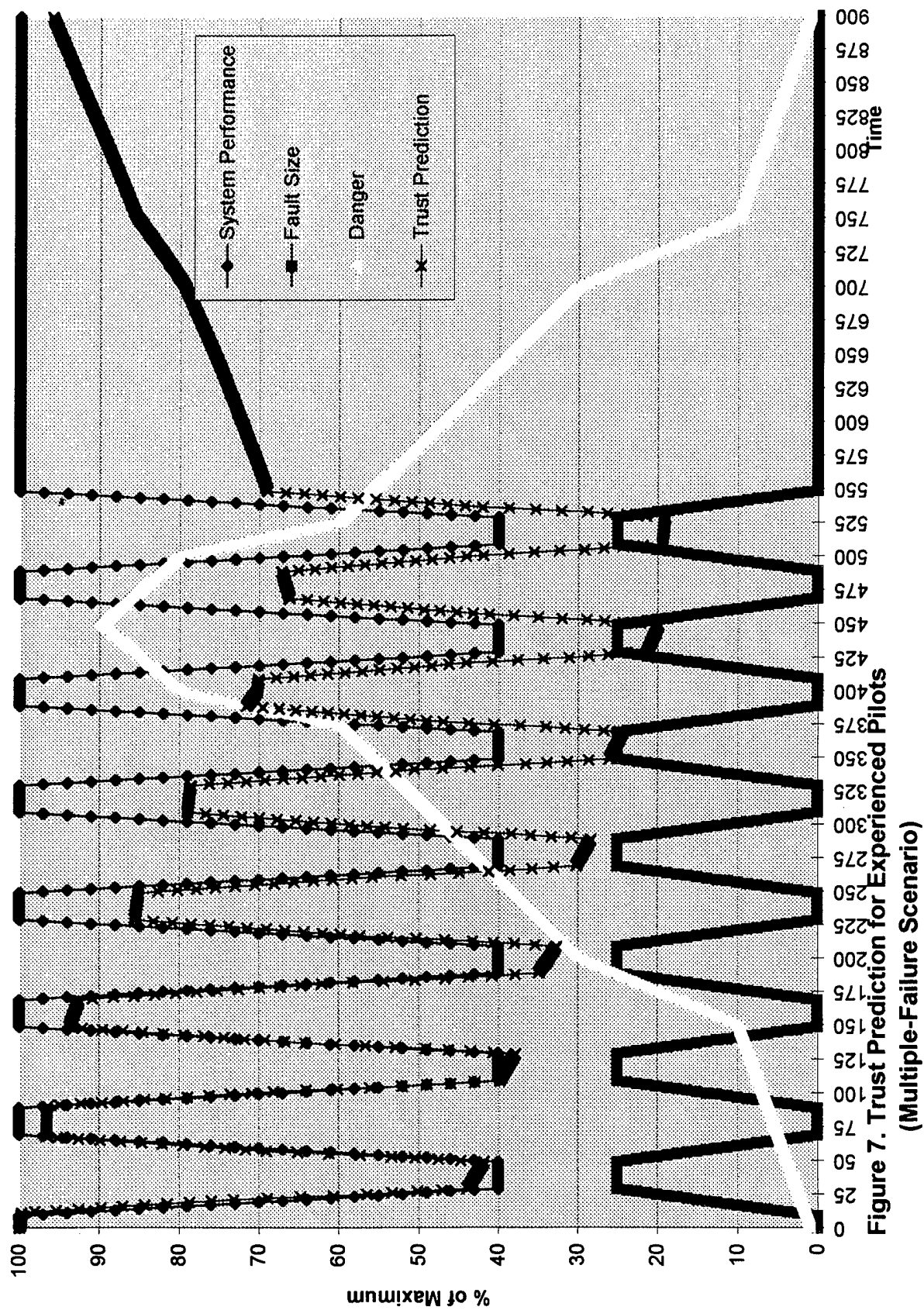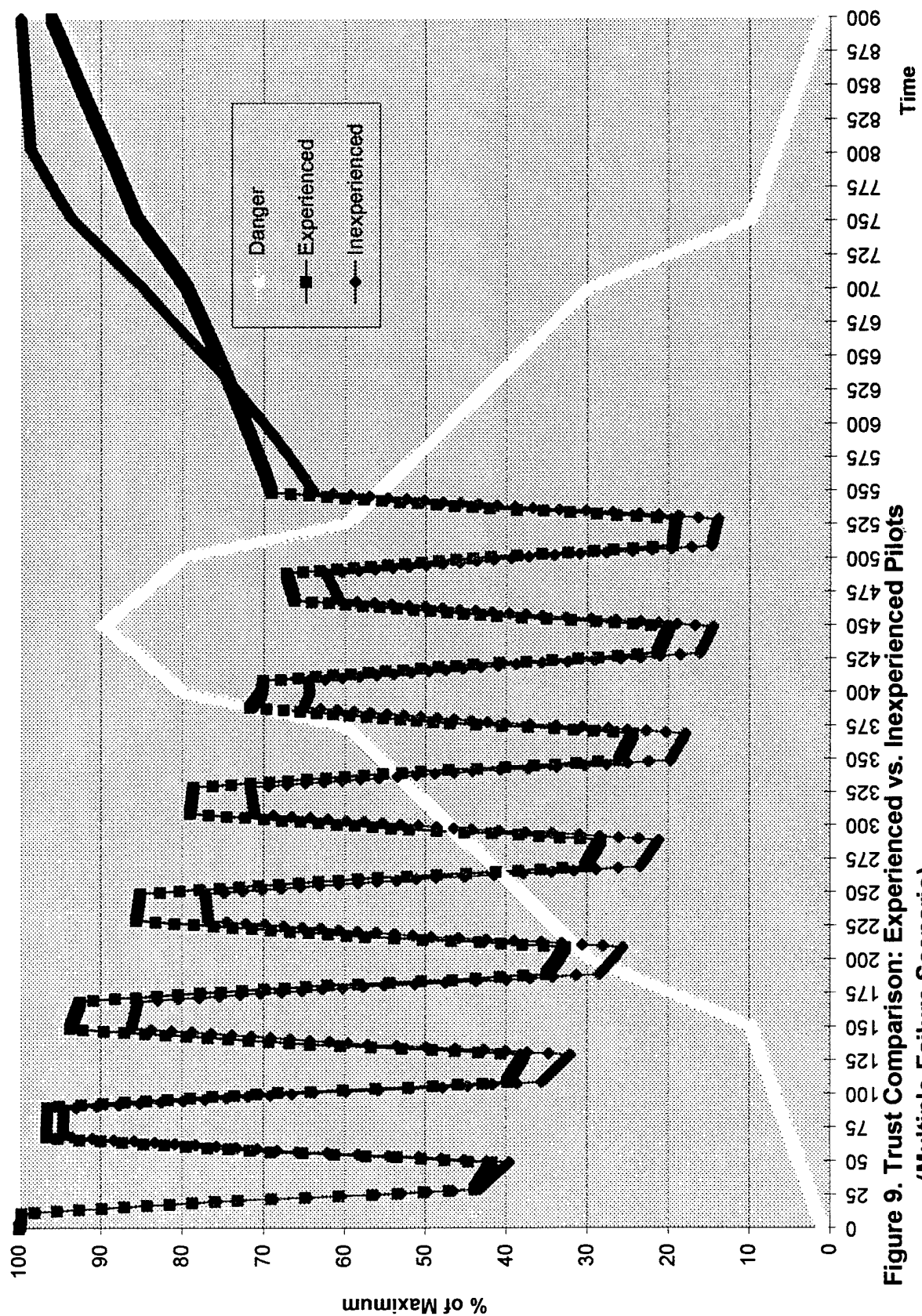Figure 9. Trust Comparison: Experienced vs. Inexperienced Pilots (Multiple-Failure Scenario)

This page left blank

7.0 THE WORKLOAD ALLOCATION MODEL

In addition to developing an equation of trust, Lee and Moray [1994] also developed an equation of workload allocation. This equation uses the level of trust as one parameter and further incorporates the pilot's personality. Their original equation is stated as:

$$\%Automatic(t) = \Phi_1 [\%Automatic(t-1)] + A_1 [\{T-S_c\}(t)] +$$

$$A_2 [IndividualBias] + A_3 [a(t)]$$

Where: %Automatic is the percentage of the workload the pilot is likely to allocate to the computer

Φ_1 is a proportion that tells how strongly the pilot's current use of automation depends on the past use of automation

T is the percent trust the pilot has in the system, computed as described previously

S_c is the percentage of self-confidence the pilot has

IndividualBias is the pilot's percentage acceptance of automation in general

a is a percentage chosen randomly from a uniform distribution, it accounts for unknown random events that may impact workload allocation

t refers to the current time

t-1 refers to the just-past time

A_1, A_2, A_3 are proportions that determine the importance of a given variable, these proportions plus Φ_1 sum to 1.0

The authors made two modifications to the original workload allocation model. The first is to interpret t-1 as a weighted moving average of the past values. Section 4.1 describes the weighting schemes. The second modification is to assign the quantity (100.0 - Danger) to S_c. In other words, S_c is the self-confidence the pilot has in "his" ability to accomplish the task within the required timeframe given the level of danger. This has the affect of decreasing self-confidence as the danger increases. (It is possible that, with some pilots, the presence of danger will increase self-confidence. We speculate on this in the last section of this report.)

8.0 TESTING THE WORKLOAD ALLOCATION MODEL

The same air-to-ground scenario described in Section 2.0, with
dual-failure and multiple-failure examples, was used to test the
workload allocation model. Two tests were run on each example.
These tests used the following proportion values to reflect the
experienced and inexperienced pilots.

		TEST 1 (Experienced Pilot)	TEST 2 (Inexperienced Pilot)
Φ_1	=	0.10	0.10
A_1	=	0.80	0.33
A_2	=	0.10	0.57
A_3	=	0.00	0.00

Note that random effects were not considered (weight A_3 = 0.0).
Test 1 reflects a pilot who places great weight on self-confidence
and the current level of trust (A_1 = 0.80). The pilot in Test 2
does not weight self-confidence as highly (A_1 = 0.33). "He" is
driven more by "his" attitude toward automation (A_2 = 0.57) because
of a lack of sureness in "his" abilities. Neither pilot has a
strong bias for or against automation (the IndividualBias term was
set to 50%).

8.1 DUAL-FAILURE SCENARIO WORKLOAD ALLOCATION PREDICTION

The test results are illustrated in Figures 10 and 11 with
comparisons between experienced and inexperienced pilots given in
Figure 12.

Figure 10. Workload Allocation Prediction for Experienced Pilots (Dual-Failure Scenario)

42

Figure{} Workload Allocation Prediction for Inexperienced Pilots (Dual-Failure Scenario)

43

Figure 12. Workload Allocation Comparison:
Experienced vs. Inexperienced Pilots (Dual-Failure Scenario)

44

The greater confidence of the experienced pilot caused "him" to take over the system manually in the case of failure. During the first failure, 29% of the workload was allocated to the automation, whereas during the second failure, only 14% was allocated. Conversely, the inexperienced pilot does not have the confidence to take over manual control to the same degree. During the first failure, the allocation was 41%. During the second failure, the allocation was 34%. The greater weight on trust permits the experienced pilot to allocate more workload to the system than does the inexperience pilot when there is high danger but no failures (67% vs. 57% respectively at time index 450). In the case of low danger and good system operation, the inexperienced pilot allocates more workload to the automation than does the experienced pilot. (Note for example time period 560-700.)

8.2 MULTIPLE-FAILURE SCENARIO WORKLOAD ALLOCATION PREDICTION

These results are illustrated in Figures 13 and 14 with comparisons between experienced and inexperienced pilots given in Figure 15.

During multiple failures, the experienced pilot allocates little if any workload, except in cases of high danger (time period 400-500). The inexperienced pilot does not have the confidence to completely take over manual control. "He" has an equal to or higher allocation throughout the mission compared to the experienced pilot. Here again, the recovery of allocation is slower after subsequent failures.
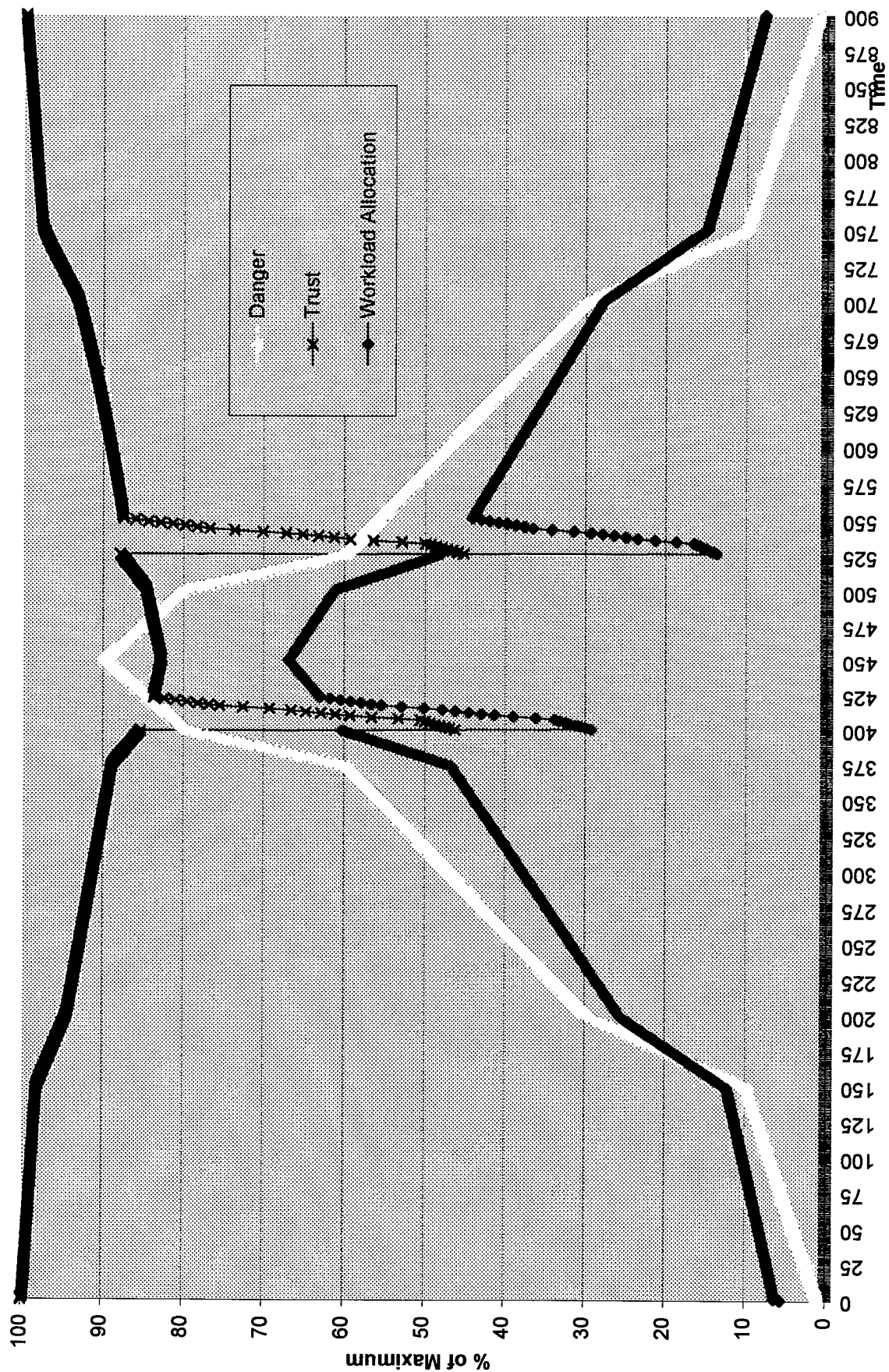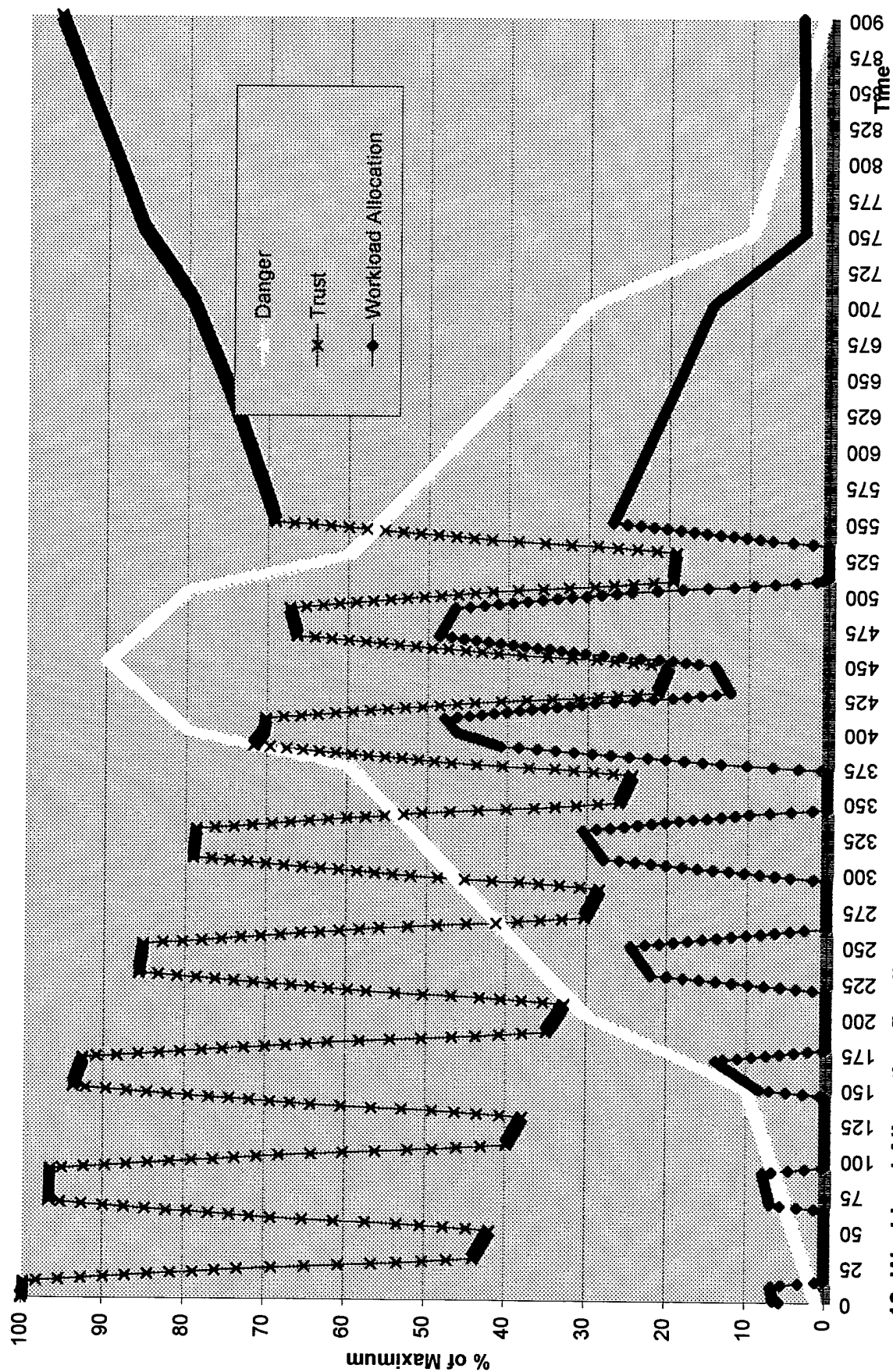
Figure 13. Workload Allocation Prediction for Experienced Pilots (Multiple-Failure Scenario)
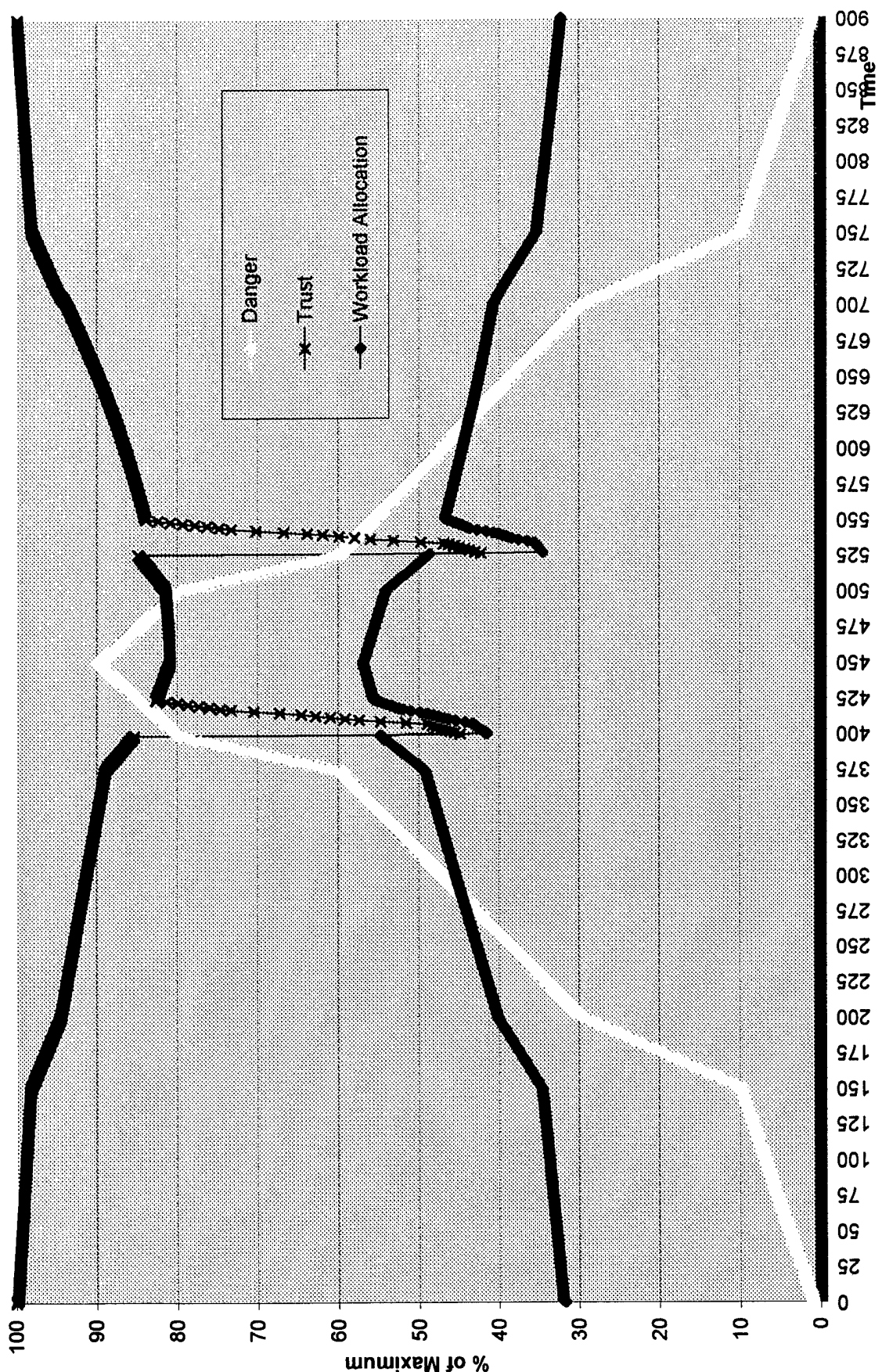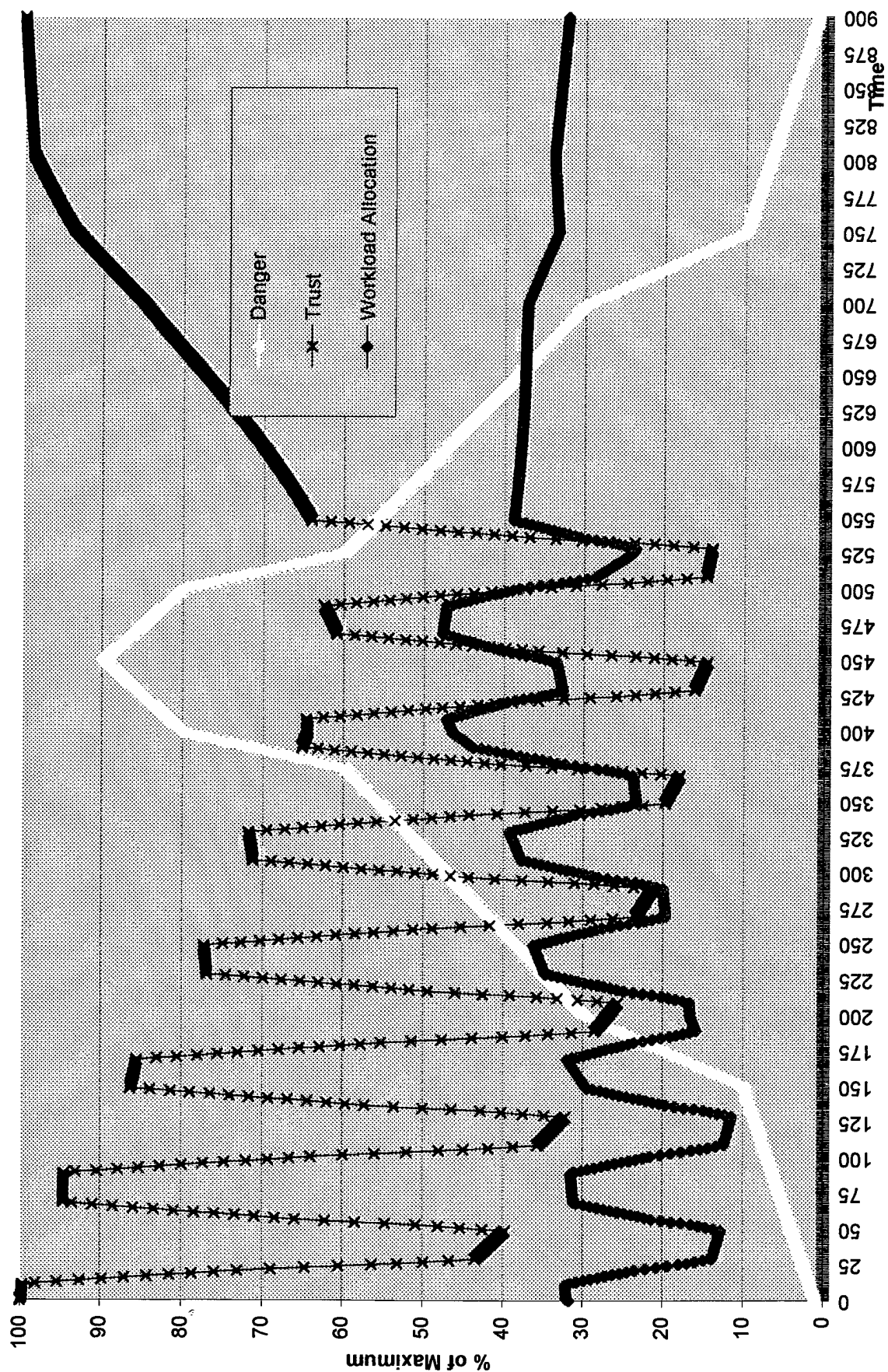
Figure 14. Workload Allocation Prediction for Inexperienced Pilots
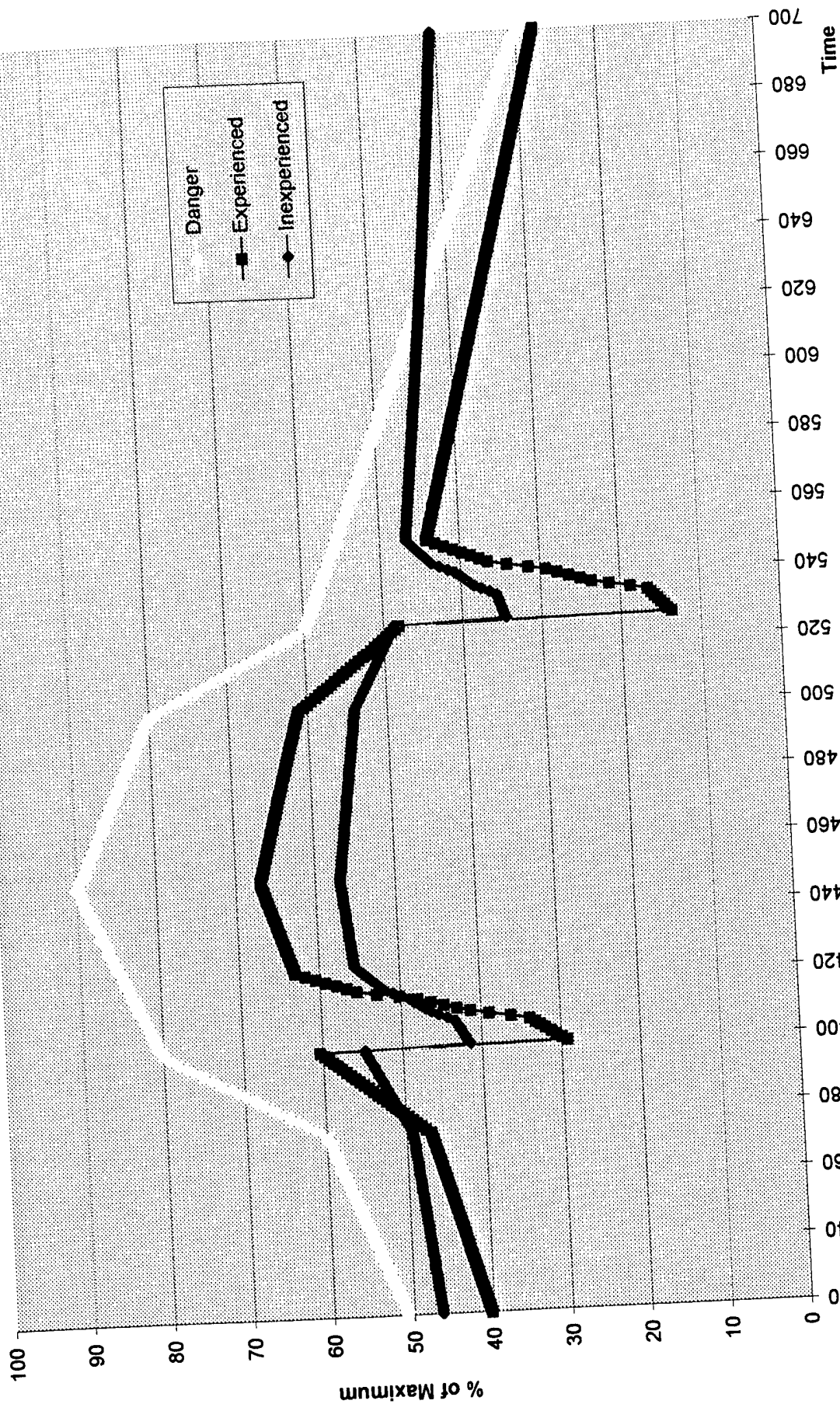(Multiple-Failure Scenario)
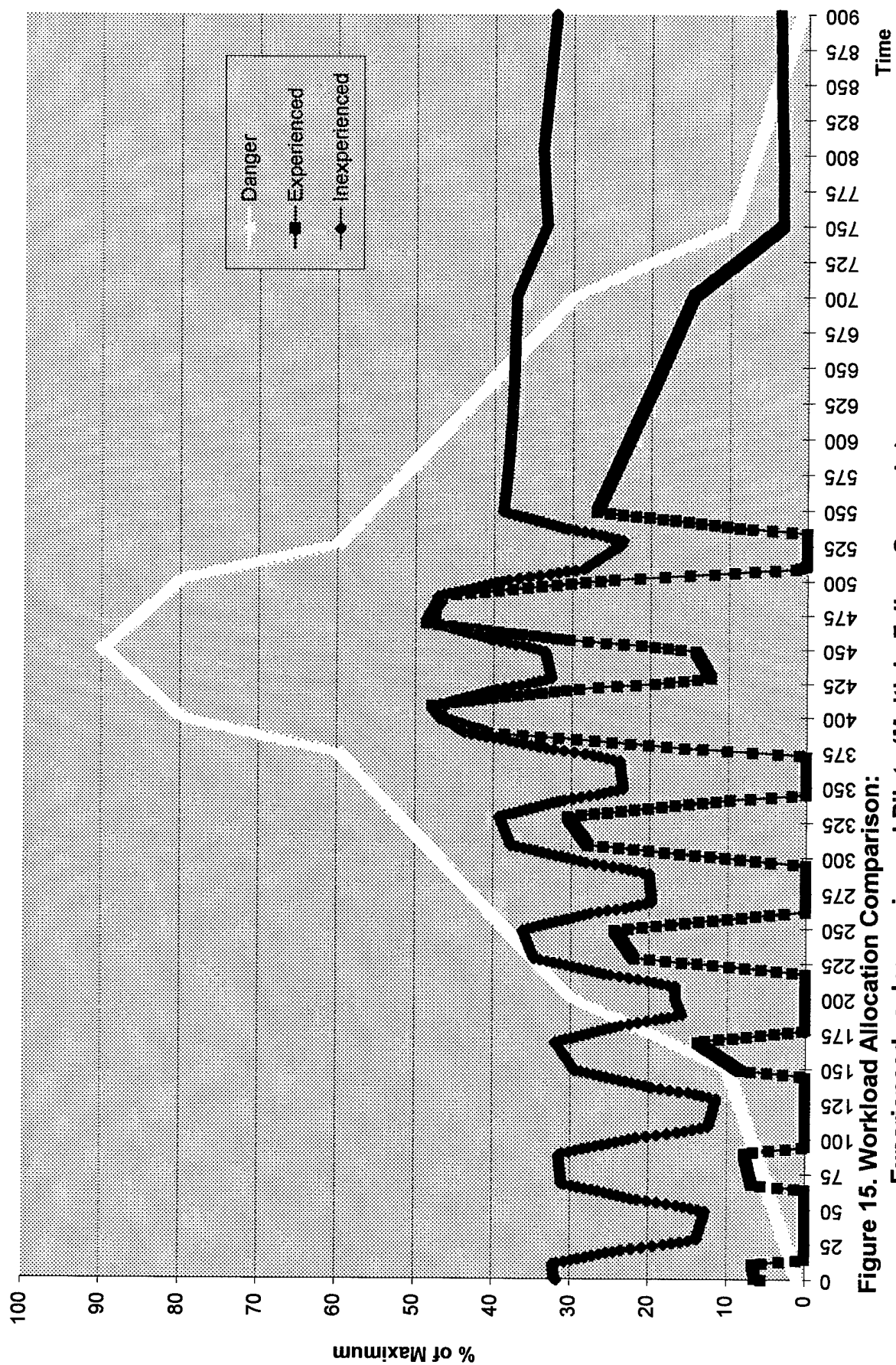
Figure 15. Workload Allocation Comparison:
Experienced vs. Inexperienced Pilots (Multiple-Failure Scenario)

48

9.0 CLOSING DISCUSSION

In this section, we will try to relate the demonstrated performance of the model described in this report to what is known about human trust. We will also offer some suggested improvements to the model. Finally, we explain the origin of the model's name, LeM^2*R^3.

9.1 RELATIONSHIP TO CURRENT THEORY ON HUMAN TRUST

Brasher performed a literature review to draw some generalizations about human-human trust that could relate to human-automation trust. His review is included in Appendix D. According to Brasher(p2), there is very little documented research on interpersonal trust that relates to human-automation trust. There is, however, sufficient material for a general discussion and some helpful pointers.

The literature gives many definitions for the word "trust". This leads to some confusion on the subject in that there is not a common baseline for discussion. The best definition is given by Muir who extends human-human trust to the area of human-computer trust. She says that trust is based on three factors: persistence, technically competent behavior, and fiduciary responsibility. We have found a considerable relationship between her definition and the model's performance.

> - PERSISTENCE is the expectation of constancy. Constancy allows people to understand and create mental models of physical processes, and to use these models to control the process and predict future system states. For the pilot,

this refers to the system operating in the same way all the time. The automation should not surprise the pilot. The model demonstrates the loss of trust and allocation when failures prevent consistency in system operation, especially during repetitive system faults.

- TECHNICAL COMPETENCE refers to expert knowledge, technical facility, and correct performance. In other words, the holding of special knowledge and the use of that knowledge to carry out tasks correctly. Whatever tasks the computer carries out, it should do so correctly and make use of all available information. The model's prediction of trust and allocation loss when system performance degrades may be seen as a growing sense of the automation's lack of competence. The model accommodates loss of system performance even when faults are not readily evident.

- FIDUCIARY RESPONSIBILITY is the obligation to actually carry out tasks and fulfill responsibilities, especially when being the sole holder of certain information or skills. This factor has to do with the <u>willingness</u> to perform a task vs. the <u>ability</u> to perform that task. This part of the definition of trust indicates that automation should work in synergy with the pilot and perform tasks as requested by the pilot or as defined in some task allocation scheme, whether static or dynamic. That the automation should work in synergy with the pilot and know when to carry out a given portion of the workload is the point of this model's prediction of allocation.

The literature (Brasher, p6) confirms that trust is the foundation of any relationship. Relationships can not move forward without it. One might expect then that a pilot will use the automation only so far as "he" feels it can be trusted. Thus, creating trust-worthy automation has to be the goal of the system designer. It is insufficient to, for instance, create a system that exhibits only technical competence. In this model, when all other conditions are constant, allocation (the pilot's use of the automation) decreases when trust decreases.

50

Brasher(pp6-7) also found support for the idea of "external contingencies" affecting the ability to trust. Danger, is an external contingency included in the model. The allocation by the pilot of workload to the computer is assumed to be directly related to danger. [4] Workload allocation is not permanent if all other factors remain constant and the level of danger changes. This is essential in the face of failures or loss of system performance. In the future, this contingency might be expanded to say that system performance will be lower, in the mission scenario depicted, if the pilot does not use the automation at some point. Currently, the model does not drive system performance as a factor of allocated workload and the current mission circumstances.

---

[4] Remember, in the model's current formulation, that self-confidence decreases as danger increases. As self-confidence decreases, the pilot will tend to allocate more workload to the computer. Note that, due to the need to survive, workload tends to be at its highest when danger is at its peak. This is especially so as the number of aircrew decreases for the same mission. For instance, conducting the entire F-15E mission with a single-seat fighter would mean reaching 160% workload for the pilot.

9.2 POTENTIAL IMPROVEMENTS TO THE MODEL

An improvement to the model might be achieved if separate weights are used for trust and self-confidence. This might permit more detailed control over the model and a more refined expression of personality. The model assumes that trust adds allocation and self-confidence subtracts allocation. Since both are currently weighted the same, the model assumes that trust and self-confidence balance each other. This might not always be the case.

An anecdotal story leads us to speculate on another improvement to the model. The authors witnessed an interview with a Hollywood stunt person. In comparing himself to the average person, the interviewee said that most people feel calm and "normal" when they, for instance, are watching television. These same people would feel exceptionally nervous if they had to jump off a cliff and depend on a small quick-release parachute for a safe landing. On the other hand, according to the stunt person, he never feels really calm or "normal" unless he is engaged in a dangerous stunt. In thinking about this interview, the authors realized that another improvement to the model might be to incorporate a variable slope and baseline into the curve governing self-confidence. The new equation of self-confidence that would result is given below. Since this speculation comes only from one anecdotal interview, no changes are currently planned based on it.

$$S_c(t) = S_{cb} + (D_f * D(t))$$

Where:

    S_c is the resulting self-confidence, this number is restricted to values between 0 and 100

    S_{cb} is the self-confidence baseline, the amount of self-confidence the pilot has when there is no danger, this number is restricted to values between 0 and 100

    D_f is the danger factor, a value between 0 and +/- 1.0, the impact that danger has on self-confidence; if this is a positive number then self-confidence is directly related to danger, if this is a negative number then self-confidence is inversely related to danger

    D is the amount of danger itself and is restricted to values between 0 and 100

    t refers to the current time

Some thought also needs to be given to the notion that, at some point during the mission, the workload may radically increase. In fact, workload may increase beyond the pilot's cognitive limits. This would force at least some workload to be allocated to the automation lest system performance degrade. A minimum allocation setting greater than 0% would account for an aircraft that, even in the best of circumstances, can not be flown without computer assistance. Beyond that, a feedback loop in the model could seek an optimum level of allocation such that allocation is increased in order to improve system performance. This is a fundamental shift in the current use of the model as a predictor of trust and allocation as a result of circumstances. This might, however, be an interesting path to follow in developing a real-time recommendation on what percentage of the workload to allocate to the automation, especially if combined with research on human cognitive limits. To make this work, some thought would also have to be given to the

setting of system performance. Currently, system performance is randomly set as an event input to the model. It would be necessary to expand this thinking to make system performance a function of workload, allocation, and mission difficulty. All three of these vary over time. Mission difficulty could be a direct function of danger. Riley [1994] proposed a model in which workload, risk, and trust are not necessarily interrelated. His model needs to be examined more closely in light of Lee and Moray's work and Muir's theory. From this examination, we can determine its best impact on the model discussed in this report.

It is possible that the rate of trust recovery would degrade with each subsequent failure relative to the time between each failure. Additionally, Riley's [1994] results showing that loss of trust does not occur with the first failure should be incorporated. At the present time, the model does not incorporate this thought but its implementation may be necessary in the future to accurately reflect pilot reaction.

Since his original work, Moray [1997] has suggested that it is not the danger itself that affects pilot trust but the perception of danger. For instance, "... there may be a tendency to relax on the way out of the danger zone. Note that many mountain climbing accidents happen on the way down, when the maximum danger is subjectively felt to be past." Additional support for this idea comes from traffic accident statistics. Some states report that the

most accidents happen within 25 miles of home. From this point of view, the danger actually increases when the pilot becomes overly relaxed, overly confident, and in a hurry to get home. The model could accommodate this idea by changing the scheme for generating self-confidence so that nearness to home and mission phase were factors. Another approach would be to modify the input to the model such that the danger factor reflects the pilot's impressions rather than what actually exists.

Riley's experiment #3 [1994] explored the impact of individual biases on allocation levels. In this first attempt at an engineering model for real-time application we did not get to incorporate his results. We are eager to apply those results as we improve the model's realism and applicability to military combat scenarios.

9.3 ADDITIONAL WORK USING THE PRESENT MODEL

The review of this work brought to light several things that might be considered to draw additional insights from the current model and to add to the model's fidelity.

According to Lee [1997], his original trust model was derived from a process using 6-minute time intervals. The scenario used in this report is based on 0.2-minute time intervals. The results would be more accurate if the scenario were based on the same time interval used to derive the original model.

Some Monte Carlo runs need to be made on the model's results to determine if there is a statistically detectable difference between the experienced and inexperienced pilots. This is especially needs to be done for trials involving few faults. This analysis needs to be coupled to an operational analysis of the differences to assess their impact from a practical point of view.

The literature on human decision-making indicates that people tend to underestimate the possibility of extremely adverse events and overestimate the possibility of extremely positive events. This is one of several human decision-making biases that is not taken into account by the model.

It may be that all types of faults do not have the same impact on trust, in part because it is unlikely that all types of faults would be detected with equal accuracy by the pilot. Similarly, the model assumes that pilots will be perfectly accurate detectors of fault magnitude and react accordingly.

According to Figure 9, even after seven faults pilots would still have over 10% trust in the automation. This may not be realistic. There may be some point at which a pilot's trust would go to 0% (equivalent to the pilot "writing off" the automation as unacceptably unreliable). Such a severe degradation would affect the pilot's recovery of trust. For example, if a pilot writes off the automation as unacceptably unreliable, trust recovery, if it

occurs at all, may be more prolonged. Riley's work contains some indication that trust may not ever recover completely.

Given that <u>perceived</u> danger is the key parameter, a pilot with poor situation awareness might not perceive a hazard and therefore the danger associated with this hazard would not affect his level of trust. Also, it seems likely that there might be a synergistic effect between automation failure and danger: the more automation faults a pilot perceives, the more danger he might <u>perceive</u>. Currently, the model only considers danger external to the aircraft, as perceived by the pilot.

9.4 ORIGIN OF THE LeM^2*R^3 NAME

The reader may wonder where the name "LeM^2*R^3" comes from. Lee and Moray performed the original experiments and produced a preliminary model. Muir developed the human/automation trust theory to guide additional work. Raeth worked from the computer engineering viewpoint to translate the theory and empirical observations into a model that could be tested in a cockpit emulation system. Reising provided the human factors and cockpit integration perspectives as a means of refining the engineering model. Finally, Riley brought in the idea of danger and variable recovery heights. He also independently confirmed many of the results of Lee and Moray.

10.0 SUMMARY

For pilots to accept active decision aids during complex flight scenarios, it is essential that the automation work in synergy with the aircrew. To accomplish this, the automation must go well beyond menu and macro selections, where the pilot must explicitly tell the automation what to do and when to do it. It must also transcend "mother may I" approaches, where the automation asks for permission to proceed. To breach these traditional barriers, the automation needs a sense of how the pilot will react in a given situation and, based on that reaction, how much of the workload could be allocated to the automation at any given time. For this purpose, the authors developed and tested a dynamic model of pilot trust and workload allocation. This "full degrees of freedom" engineering model transitions theoretical and empirical observations published by several experts. It can be combined with other information obtained from static and continuous processes to divide the workload and minimize cognitive overload.

Appendix A

GENERATING WEIGHTED MOVING AVERAGES

In this appendix we derive the equations to generate weights based on moving averages.

A.1 WEIGHTED MOVING AVERAGES. According to Kazmier, the weighted moving average is computed as follows:

$$\overline{X_w} = \frac{\sum_{t=1}^{n} w_t X_t}{\sum_{t=1}^{n} w_t}$$

Where:

$\overline{X_w}$ is the weighted average
X_t is a discrete value within the average
w_t is the weight or importance of a discrete value
    (it is necessary for this value to be $>= 0.0$)
t represents time incrementing by 1. In the case of
    calculating a moving average, the history starts at the
    current time sample minus n. Therefore, if the current
    time stamp is 22 and n is 5, the moving average is
    calculated from time stamps 17 through 21.
n is the integer number of discrete values
    (The response speed of the moving average curve to the
    original data is inversely proportional to n. As n
    increases, small perturbations in the original data are
    filtered out.)

A weighted moving average is developed by moving the n-sized window incrementally down a series of discrete values, from the earliest to the most recent. The total number of discrete values is assumed to be $>= n$.

59

A.2 WEIGHTS FOR MOVING AVERAGES. At issue with a weighted moving average is how to develop the weights on the individual values. At first, we simply made all the weights equal to each other. This gave us the effect of a non-weighted average. The non-weighted average assumes that a pilot will be affected by the relatively distant past as much as by the most recent present. (This is not a realistic assumption.) We then sought a means of weighting the most recent values in the moving average several times more heavily than the least recent values and having the weights decrease between those two extremes.

The following table illustrates this idea:

```
                                          | |  Most Recent
                                          \/  Present
TIME      1    2    3    4    5    6    7    8    9    10

VALUE     6    10   13   14   15   16   15   14   12   11

WEIGHT    1    1    2    2    3    3    4    5    6    7
```

A.3 THREE WEIGHTING METHODS. Given the large amount of sample data to be generated, an equation is needed to automatically develop the weights. Three weight generation equations were made available to this model (we used only the last two): constant value, linear decreasing, and sigmoid decreasing. Any of the equations can be chosen for each parameter prior to running the model. These three equations are illustrated in Figure 16.
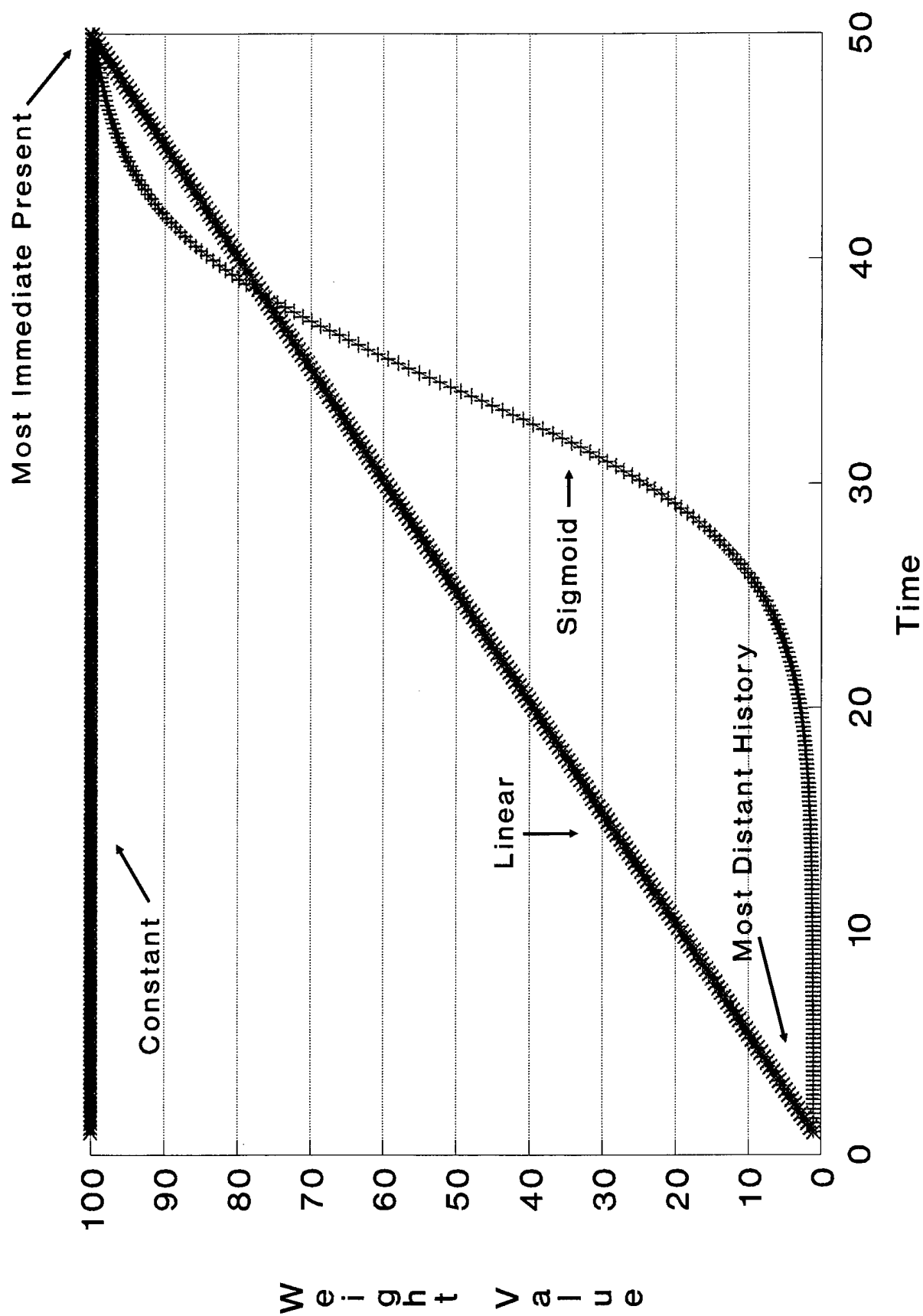
Figure 16. Results From Weight Equations

The parameters used to generate the equations were chosen to give sufficient difference in the shape of the curves, as a means to illustrate the reactions of different pilots. Empirical experiments and additional consideration of human trust theory are needed to precisely determine the value of these parameters. Each equation has its unique affect on the view of history taken by the pilot. To illustrate, assume a decision aid fault causes a severe loss of system performance. However, the system's performance has been perfect up until the time of the fault. The pilot's view of the system's historical performance will be different depending on the weighting scheme used.

Constant weighting uses a non-changing value as the weight. It places the same importance on all values regardless of when they occur. (This scheme is also called "unweighted" since the effect of the constants is mathematically canceled.) An example of the effect from constant weighting appears in Figure 17. The moving average shows the pilot's interpretation of system performance history when the samples are all weighted equally. Note that, after the system fault occurs, the reaction of the history curve to the loss of performance is not as quick nor as severe as in the linear or sigmoid cases. The moving average only drops from 95 to 90. Once the failure has been cleared, recovery does not take place within the time period shown. This type of average shows general trends in the data: the longer the time window's length (n), the more general the trend.

Linear weighting uses weights whose values change at a given rate. For our purposes, lower weights are attached to earlier samples. Thus, the pilot places more importance on what occurred recently. An example of the effect from linear weighting is shown in Figure 18. Note that the reaction of the history curve to the fault is more extreme compared to the constant-weight scheme but less so compared to sigmoid weighting. The moving average drops from 95 to 85. Some slight recovery does occur within the time window shown (to 87). The details of generating linear weights are discussed below.

Sigmoid weighting is one means to make weights much greater for more recent values than for older values. The effect is to place an exponentially increasing importance on recent occurrences. An example of sigmoidal weighting is shown in Figure 19. Note that the response of the history curve to the fault is the most extreme of the three methods. The moving average drops from 95 to 76. This is to be expected since the weights are so much higher for more recent values than is true for constant or linear weighting. The recovery rate is the highest of the three methods once the fault has been cleared, bringing the curve back to 80 within the time window. The details of generating sigmoidal weights are discussed below.

For both the linear and sigmoid weighting schemes, the more experienced the pilot, the shallower the slope of the curve. Extreme differences between experienced and inexperienced pilots can be achieved by not only varying the slopes but also by varying the maximum and minimum values of the curves.

In the following two subsections we derive the equations that generate linear and sigmoid declining weights.

Figure 18. Linear Weight Moving Average
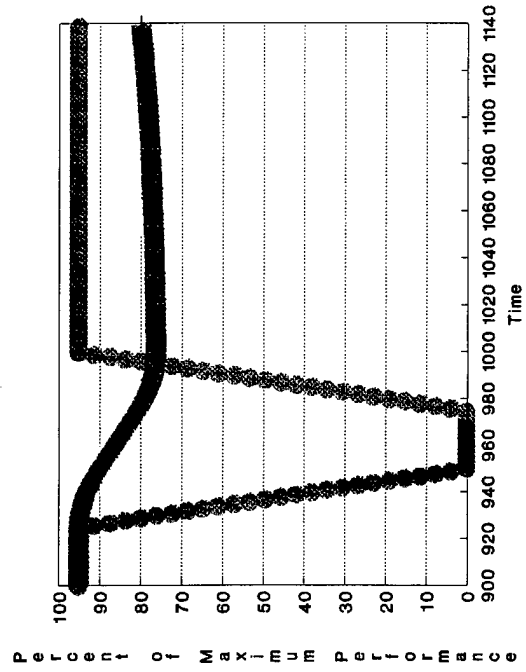
Figure 19. Sigmoid Weight Moving Average

Figure 17. Constant Weight Moving Averag

65

A.31 GENERATING LINEAR WEIGHTS. According to Wooton and Drooyan, it is a simple matter to choose a value from a line since a line is easily determined by the selection of two points. In our case, we need to choose the value of the line at $t = a$ and the value at $t = a+(n-1)$, where a is the earliest value of time.

The basic equation of a line may be stated as: $At + B = L$. To get the exact equation for a given sequence of weights, it is necessary to solve two equations in two unknowns. Let L_1 be the weight you choose for $t = a$ and let L_2 be the weight at $t = a+(n-1)$. Let $t_1 = a$ and let $t_2 = a+(n-1)$. (Note: $L_1 < L_2$ and $t_1 < t_2$) Then, the two equations can be stated as: $At_1 + B = L_1$ and $At_2 + B = L_2$. A symbolic solution of this system yields a generic equation that can be easily automated.

To accomplish the symbolic solution, subtract one equation from the other. This yields $A(t_2 - t_1) = (L_2 - L_1)$ so that $A = (L_2 - L_1) / (t_2 - t_1)$. Substituting back into the equation for L_2 yields $B = L_2 - t_2((L_2 - L_1) / (t_2 - t_1))$. Now, if $t_1 = a$ and $t_2 = a+(n-1)$, $t_2 - t_1 = a + n - 1 - a = n - 1$. So, one can simplify things by just dividing by the length of the moving average, less 1.

Given the above derivation and letting $L = w_t$, the following equation will deliver the linear weight for any point in a moving average: $w_t = t((L_2 - L_1)/(n-1)) + L_2 - t_2((L_2 - L_1)/(n-1))$.

The plot shown in Figure 18 uses linear weighting with the following parameters:

$L_1 = 1$
$L_2 = 1800$
$t_1 = 1$
$t_2 = n = 900$

A.32 GENERATING SIGMOID WEIGHTS. Calculating the sigmoid weight is somewhat more difficult. This scheme puts far more emphasis on the most recent present. According to Rogers, the basic sigmoid equation is:

$$S = E \left(\frac{D}{1 + e^{-A(t-B)}} + C \right)$$

Where:

    S is the value of the sigmoid calculation
    A makes the slope steeper or shallower
        (bigger numbers give steeper slopes)
    B shifts the curve left (-) or right (+)
    C shifts the curve up or down
    D raises and lowers the basic peak
    E raises and lowers the overall peak
    e is the base of natural logarithms

Assuming that one wants the value of the sigmoid curve to be lower at earlier periods of time, it is possible to develop a general sigmoid equation that will accommodate various beginning and ending weight values for given periods of time. (As in the linear scheme, it is possible to develop a general equation for cases where the weights are higher at earlier periods of time but that is a slightly different derivation.)

67

Similar to the discussion for linear weight equations, let S_1 be the weight choice for the earliest time (t_1) and S_2 be the weight for the most immediate present (t_2). (Note: $S_1 < S_2$ and $t_1 < t_2$) Since the sigmoid goes asymptotic to its lowest and highest values, it is necessary to define those values. Thus, multiply S_1 by 0.99 and S_2 by 1.01. In that way, it is possible to get close to a desired maximum and minimum without causing numeric under-flow or overflow. This will cause the calculated maximum and minimum weights to be slightly off but we have not found that to be a problem in the larger scope of its impact on the model's results.

Now the two equations in two unknowns become:

$$S_1 = \frac{1.01 S_2 - 0.99 S_1}{1 + e^{-A(t_1 - B)}} + 0.99 S_1$$

and

$$S_2 = \frac{1.01 S_2 - 0.99 S_1}{1 + e^{-A(t_2 - B)}} + 0.99 S_1$$

Some explanation of these two equations is in order. Since we want two equations in two unknowns, we take the simplifying step of setting E of the general sigmoid to 1. Now the minimum value of the sigmoid will be C and the maximum value will be C+D. So, for our case, $C = 0.99S_1$ and $C+D = 1.01S_2$. This being the case, $D = 1.01S_2 - 0.99S_1$.

Algebraic manipulation of the above two equations yields:

$$AB - At_1 = \ln(((1.01S_2 - 0.99S_1) / 0.01S_1) - 1) \qquad = \ln(x)$$

and

$$AB - At_2 = \ln(((1.01S_2 - 0.99S_1) / (S_2 - 0.99D)) - 1) \quad = \ln(y)$$

The procedure from here is the same as for the linear weight scheme. Subtracting the two equations yields:

$(At_2 - At_1) = \ln(x) - \ln(y)$ so that

$A = (\ln(x) - \ln(y)) / (t_2 - t_1)$. As before, $(t_2 - t_1) = (n - 1)$. Now substitute back into the second sigmoid equation and

$B = (\ln(y) + At_2) / A$.

Given the above derivation and letting $S = w_t$, the following equation will deliver the sigmoid weight for any point in a moving average:

$$W_t = \frac{1.01S_2 - 0.99S_1}{1 + e^{-A(t - B)}} + 0.99S_1$$

The plot shown in Figure 19 uses a sigmoidal weighting scheme with the following parameters:

$S_1 = 1$
$S_2 = 1800$
$t_1 = 1$
$t_2 = n = 900$

Appendix B

BIBLIOGRAPHY

Brasher, Jeffrey D. "Generalizing Trust From Humans to Automation: A Review of the Literature", Internal Report, Veda Inc, Dayton OH, Air Force Contract # F33615-93-D-3800, Delivery Order # 15, included in this report as Appendix D, Jul 95

Kazmier, Leonard J., Business Statistics, New York, New York: Mc-Graw Hill, p 29, 1976

Lee, John D. and Neville Moray "Trust, Control Strategies, and Allocation of Function in Man-Machine Systems" Ergonomics, vol 35, # 10, pp 1243-1270, Oct 92

Lee, John D. and Neville Moray, "Trust, Self-Confidence, and Operator's Adaptation to Automation", International Journal of Human-Computer Studies, vol 40, pp 153-184, 1994

Lee, John D., Personal Communication, Seattle, WA, Battelle Human Affairs Research Center, Apr 95

Lilley, Douglas, Personal Communication, Lockheed Aeronautical Systems Company, Marrietta Georgia, Jul 95

Moray, Neville, Personal Communication, 1997

Morrison, Jeff G. (Ed.) Adaptive Function Allocation for Intelligent Cockpits, Cameron Station, VA: National Technical Information Service, U.S. Navy Report # NAWCADWAR-93031-60, May 93

Muir, B.M. "Trust in Automation: Part I. Theoretical Issues in the Study of Trust and Human Intervention in Automated Systems", Ergonomics, v 37, # 11, pp 1905-1922, 1994

Pope, Alan T. and Edward H. Bogart, "Method of Encouraging Attention by Correlating Video Game Difficulty with Attention Level", NASA Technical Note # LAR-15022, United States Patent 5,377,100 dated 27 Dec 94

Raeth, Peter, G.; Noyes, James, L.; and Anthony Montecalvo "Trust-Enhancing Sensor and Information Fusion For Knowledge-Based Cockpit Decision Aids" Proceedings: USAF/RAF International Workshop on Human-Computer Teamwork, Cambridge, England, Sep 94

Raeth, Peter, G. and John M. Reising, " ", ... the NAECON paper

71

Reising, John, M., "Class Notes in Human Factors Engineering", University of Dayton, Dayton, OH, Department of Psychology, 1995

Riley, Victor A., "A General Model of Mixed-Initiative Human-Machine Systems", Proc: Human Factors Society 33rd Annual Meeting, Oct 89

Riley, Victor A., Lyall, Elizabeth, and Earl Wiener; Analytic Methods for Flight-Deck Automation Design and Evaluation, Honeywell Technical Report, FAA Contract # DTFA01-91-C-00039, Dec 93

Riley, Victor A., Human Use of Automation, Dissertation: University of Minnesota Graduate School, 1994

Riley, Victor A., Personal Communication, 1997

Rogers, Steve, "Class Notes in Neural Networks", Air Force Institute of Technology, Dayton, OH, Department of Electrical and Computer Engineering, 1990

Wooton, William and Irving Drooyan, Intermediate Algebra, Belmont, CA: Wadsworth Publishing, p 178, 1968

Appendix C

AUTHOR BIOGRAPHIES

PETER G. RAETH received the B.S. degree in Electrical Engineering from the University of South Carolina in 1979 and the M.S. in Computer Engineering from the Air Force Institute of Technology in 1980. During the period of this research, he was a Major on active duty with USAF Wright Laboratory in Dayton, Ohio and led the Vehicle-Pilot Integration Technology Team. This organization conducts, manages, and transitions research in advanced primary flight displays, information controls, and automated decision aids for operational aircraft. He has published articles, papers, and reports on the subjects of expert systems and neural networks since 1977. His book, "Expert Systems: A Software Methodology for Modern Applications", was published by the IEEE Computer Society in 1990. In 1991 he was awarded a research fellowship in neural networks at the University of Dayton Research Institute. He was named to the top 20 technical leaders in the Dayton Ohio region for 1994 and 1995 by the Dayton Affiliate Societies Council. In 1997 he departed active military duty and joined Simulation Technologies, Inc (STI) as a technical project leader. His duties included establishing the company's infrastructure for distributed parallel processing and simulation. He is currently with Ball Corporation where he transitions technology for infrared signal analysis. He is a member of Tau Beta Pi, Eta Kappa Nu, Omicron Delta Kappa, IEEE Computer Society, and the ACM Artificial Intelligence SIG.

JOHN M. REISING is an Engineering Psychologist in Wright Laboratory's Advanced Cockpit Branch located at Wright-Patterson Air Force Base Ohio. After receiving his PhD in Industrial Psychology from Southern Illinois University in 1969, Dr. Reising joined the Bunker-Ramo Corporation where he worked as a Human Factors specialist. His primary work efforts centered around the design of advanced cockpits and the execution of experiments to investigate new control and display concepts. In 1972 he left Bunker-Ramo to join what is now called the Flight Dynamics Directorate of Wright Laboratory where he is still employed. His current research still centers around advanced cockpit design, with a special emphasis on blending the many new cockpit technologies so that the pilot can use them optimally. Currently under examination are 3-D stereo cathode ray tubes, flat panel displays, touch sensitive overlays, voice controls, and programmable switches. The focus of his research revolves around developing cockpit technologies that will facilitate and maximize effective communication between the pilot of the future and the artificially intelligent computers that will be housed in tomorrow's aircraft.

He is a Wright Laboratory Fellow and a member of the Human Factors and Ergonomics Society; he was elected a Fellow within that Society in 1989. He is also an adjunct professor in the Engineering Management Department of the University of Dayton, where he has taught Human Factors Engineering since 1978. He has published over 100 papers/journal articles/technical reports.

APPENDIX D

GENERALIZING TRUST FROM HUMANS TO AUTOMATION (A REVIEW OF THE LITERATURE)

Jeffrey D. Brasher
Veda Incorporated
5200 Springfield Pike, Suite 200
Dayton, Ohio 45431-1289

31 July 1995

ACKNOWLEDGMENTS

INTRODUCTION

The introduction of advanced avionics in cockpit design provides the pilot with many enhanced capabilities. However, these new capabilities sometimes bring increased complexity and, subsequently, increased workload. In an effort to compensate for these increases, many research and design efforts have turned towards automation. Automating systems inherently changes the role of the pilot from operator to system supervisor. The goal of automation in the cockpit is the transfer the performance of tasks from the pilot to the system, thereby decreasing the pilot's workload. However, the pilot's acceptance and use of automated systems are crucial to the realization of this goal.

A key issue in the acceptance of automation is trust. In order for pilots to effectively use an automated system, he/she must acquire and maintain a degree of trust in the system. The requirement for trust raises a host of questions. What is trust? How is trust achieved? How is trust maintained? To answer these questions, we must understand the very nature of trust. It seems, however, that before we can begin to develop methods of developing and maintaining operators' trust in automation, we must first have an understanding of how trust is developed and maintained between humans.

In this paper, I will primarily review some of the research that has been reported regarding trust between humans, or interpersonal trust. It is important to note that there has been

surprisingly little research done in the area of trust,
especially since it is so central to our society. From this
research, I will attempt to relate these findings to the trust
that exists between humans and automated systems.

DEFINITIONS OF TRUST

Trust has been defined many different ways throughout the years.
Further, as noted by Barber (1 983), there is a vagueness that
exists in the multiple meanings of trust. Even Webster's Third
New International Dictionary illustrates this point:

> 1a. assumed reliance on some person or thing: a confident
> dependence on the character, ability, strength, or truth of
> someone or something
>
> 2a. dependence on something future or contingent: confident
> anticipation
>
> 5a(1). a charge or duty imposed in faith or confidence or as
> a condition of some relationship

It is important to note that all of these definitions imply an
expectation of some kind. However, the different kinds of
expectations are not clearly distinguished. The variety of
meanings is even indicated in the list of synonyms: confidence,
reliance, dependence, faith. Further, one definition of trust
may apply to a certain social relationship, while be
inappropriate for another relationship. Is it possible to
integrate the many meanings of trust into one comprehensive
definition that is applicable to all relationships? (Barber,
1983).

It is surprising that little research has been done in the area
of trust. Further, this research has resulted in even greater
ambiguities in the definition of trust. As cited by Muir (1994),
a review of the psychological literature produces the following
definitions:

- the confidence that one will find what is desired from
  another, rather than what is feared

- an actor's willingness to arrange and repose his/her
  activities on [Another] of the confidence that [Another]
  will provide expected gratification

- a generalized expectancy held by an individual that the
  word, promise, oral or written statement of another
  individual or group can be relied on

- a generalized expectation related to the subjective
  robability an individual assigns to the occurrence of some
  set of future events

- the degree of confidence you feel when you think about a
  relationship

Muir notes that, while these definitions do show several
commonalties, and illustrate the multidimensional character of
trust, they too lack the clarity that is so greatly desired
(Muir, 1994).

Barber (1 983), in recognition of his own ambiguity, proposes the
following comprehensive definition of trust:

> Our general expectation of the persistence of the natural
> physical order, the natural biological order, and the moral
> social order; our specific expectation of technically
> competent role performance from those involved with us in
> social relationships and systems; and our specific
> expectation that partners in an interaction will carry out
> their fiduciary obligations and responsibilities, that is,
> their duty in certain situations to place others' interests
> before their own (Barber, 1983).

Muir (1994) acknowledges Barber's definition as one that "explicitly recognizes the multidimensional character of trust, and at the same time includes the necessary [expectations]" identified above (Muir, 1994). Each of these expectations, persistence, technical competence, and fiduciary responsibility, is discussed below in greater detail, and extended to the area of automated systems.

Persistence is an expectation of constancy, which, according to Barber, is the foundation that trust is built upon. All humans expect that the natural and moral order of society will persist and be realized. It is exemplified in statements like, "I trust the heavens will not fall," and "I trust my fellow man to be good, kind, and decent." These are the expectations that are necessary for effective human interaction to continue (Barber, 1983). Muir (1994) extends persistence to the area of trust in automation. It is the constancy of physical laws that allows us to understand and create mental models of physical processes. It also allows us to use these models to control a physical process and predict future system states (Muir, 1994).

The expectation of technically competent behavior can refer to three classifications: expert knowledge, technical facility, or everyday routine performance. It is what is meant when people say something like, "I trust my doctor to perform the operation well" (Barber, 1983). The expectation of technical competency is probably the closest to our understanding of trust in automation

in that automated systems are designed to perform a task, and we expect them to work properly (Muir, 1994).

Finally, fiduciary responsibility is the obligation that people have to one another carry out their tasks and responsibilities in a way that demonstrates an interest above their own. It is usually placed on an individual or group based upon some special skill or knowledge they possess. And, is usually instilled to individuals who have a certain degree of power, whether it be parents, government officials, or professionals (Barber, 1983). Fiduciary responsibility may be related to automation in that there can be times that the system has information, authority, and/or power that the system operator does not (Muir, 1994).

However, it is important to realize that to gain a complete understanding of trust, if that is at all possible, it is not enough to merely define trust. We must also understand how trust is acquired, and subsequently maintained.

THE DEVELOPMENT AND MAINTENANCE OF TRUST

The development of trust has been recognized as an initial step in forming healthy relationships. Erikson (1963), in his stages of psychosocial development, identifies trust versus mistrust as the first obstacle that must be overcome. The unsuccessful resolution of this stage has direct implications on the ability to resolve subsequent stages of development (Erikson, 1963). Similarly, the inadequate development of trusting behavior has

direct implications on our ability to form healthy relationships.

The question, however, is that of how trust develops. One explanation is that it develops naturally from the contingencies inherent in a developing social relationship. One such contingency, as observed by Olvera and Hake (1976) in a study based upon social exchange theory, is the adverseness of side effects, such as aggression and lack of cooperation, that may occur between people as the result of mistrust. However, when only these natural contingencies were present, subjects exhibited only minimal trusting behaviors (Olvera and Hake, 1976). Other research has shown that external contingencies may be necessary in order to elicit trusting behaviors. In one such study, Matthews (1977) reports that trust develops only after a punishment contingency, in the form of loss of money, was introduced. In other words, trusting behaviors were only observed when reinforced (Matthews, 1977). An important observation here is that these contingencies may have a converse effect as well. Matthews, Kordonski, and Shimoff (1983) introduced temptation as an external contingency not to trust. Significant levels of mistrust were observed (Matthews, Kordonski, and Shimoff, 1983). Similarly, we might conclude that internal contingencies not to trust may also exist.

Hake and Schmid (1981) report similar findings in their study of a two-person social exchange situation. In this study, pairs of subjects were given monetary reinforcers for correctly solving

simple problems. The desired outcome of each session, as with any social exchange situation, was an equal amount of reinforcers. Minimal trusting behaviors were characterized by strictly alternating the receipt of reinforcers between the two partners. Expanded levels of trust were observed when one subject allowed his/her partner to greatly exceed an equitable amount of reinforcers with the understanding, or trust, that he/she would be allowed to catch up before the end of each trial. When no external contingencies were introduced, i.e., only natural contingencies existed, only minimal trusting behaviors were observed. Expansion of trust beyond minimal levels was only observed after an external contingency, the opportunity to earn more money, was introduced. However, these observations are limited only to the development of trust. Upon the removal of the external contingency, the acquired level of trust endured. Hake and Schmid attribute this endurance to the notion that once individuals have learned that inequity is only temporary, external contingencies are not necessary: "as long as equity is reached, what difference does it make who is temporarily ahead or behind" (Hake and Schmid, 1981).

The findings of these types of studies may have some useful implications in the area of trust in automation. Clearly, we cannot simply provide monetary reinforcement to operators when trusting behaviors are exhibited. Perhaps there is some type of schedule that may be employed so that when trust in the system leads to a successful outcome (i.e., cost-effective operation),

the operator's behavior may be reinforced. Similarly, when mistrust in the system, leading to the operator assuming control, results in unsuccessful outcomes (i.e., human error, costly down-time), a negative reinforcer should be employed. Further, as observed by Hake and Schmid, we may conclude that once the operator develops a history with the system, appropriate levels of trust will be maintained.

Other research has demonstrated that the development or restoration of trust may be achieved through a clinical approach. Greben (1984) describes the use of psychotherapy in establishing trust between the patient and the therapist. From that foundation, the patient may then be able to reestablish trust with others (Greben, 1984). Mitchell (1990), using a counseling approach, outlines topics that may be discussed that will provide the student with information concerning the negative effects of mistrust and provide some rationale for trusting. These topics may include:

- interpersonal trust as it relates to psychosocial competence; without trust, the individual has low-self esteem and feels isolated and already betrayed this sense of rejection could become progressive and even lead to paranoia

- aloneness may limit an individual in what he/she can accomplish

- no major or enduring relationship can exist happily and comfortably without trust

Mitchell then suggests some tasks the student may complete in order to begin developing trust behaviors. These include:

- keep a daily recording of examples of others' trustworthiness, paying attention to specific interactions and behaviors that may have promoted trustworthy behavior

- review past disappointments in which the student has recovered

- begin working with low risk issues, gradually working towards higher risk areas

- develop goals that are observable, measurable, and achievable

- complete a review of specific, observable changes in trusting behaviors (Mitchell, 1990)

These studies suggest that interpersonal trust is something that can be fostered or taught. These techniques may also have some use in the trust of automated systems. If trust can be taught or trained, perhaps there are also training techniques that may be used to train operators to trust automated systems. Provide the operator with rationale for trusting. Examples may include:

- Lack of trust leads to increased workload

- Description of the errors that can occur as a result of mistrust and the subsequent assumption of control

Further, there may be tasks that the operator can complete to increase trust in the system:

- keep a log of the system's behaviors

- develop trust in stages, begin with low risk areas and spread to higher risk areas

- develop goals that are observable, measurable, and achievable

These rationale and tasks should provide the operator with a strong understanding of the system's behavior, and foster trusting behaviors in the operator.

Other studies report findings that are relative to physical attributes. Brownlow and Zebrowitz (1990) report that people who are babyfaced, less aged, and generally more attractive were perceived as being more trustworthy. Further, individuals who smile were also judged to be more trustworthy (Brownlow and Zebrowitz, 1990). Darby and Jeffers (1988) report the same findings with regard to attractiveness (Darby and Jeffers, 1988). Other findings are that people who talk fast were perceived as being less trustworthy (Woodall and Burgoon, 1983), and that defendants of crimes who exhibit symptoms of sympathy, rather than guilt or no emotions, were judged as more trustworthy (Frank, 1992). Additionally, there are studies reporting differences relative to gender (Jeanquart, 1992; Harper, 1993; & Heretick, 1981) and ethnicity (Lagace & Gassenheimer, 1989; Switkin & Gynther, 1974; & Terrell & Barreft, 1979).

While these may be interesting, they do not seem to have strong applications in the area of automation. One might conclude that automated systems should be kept in good appearance, and that displays should be as simple as possible. Dilapidated and unattractive systems, or systems that have complex displays may be judged as less trustworthy by the operator.

One area of interest that was expected to be found in the literature is that of what factors cause changes in trust over time. This, however, is not the case. Research involving the development of trust over time has reported only correlational findings (Kaplan, 1973; Rempel, Holmes, & Zanna, 1985). These studies indicate that the success of a relationship and the degree of trust exhibited is positively correlated, providing the history of those involved is good. From that, we might conclude that the opposite, a negative correlation, would be realized if a poor history exists between those involved in the relationship. These findings are hardly surprising, and are also expected to hold true in the area of trust in automation.

CONCLUSIONS

The research described in this paper reports techniques for which
trusting behaviors may be developed, maintained, and/or restored
through behavioral and cognitive approaches. These include:

- provide the pilot with a rationale for trusting

- keep a log of the system's behaviors

- develop trust in stages

- develop observable, measurable, and achievable goals

While the manipulation of external contingencies may be more
difficult than the training approach described above, there may
still be some useful knowledge gained from these types of
studies:

- be cognizant of any internal or external contingencies
  that may be present

- if possible, lessen the impact of any contingencies that
  may have adverse effects

Additionally, there may even be some lessons learned from
examining the definition of trust:

- systems should perform consistently, allowing pilots to
  predict future system states (i.e., the system performs
  reliably)

- systems should perform competently (i.e., outcomes should
  be valid)

- information and authority should be carefully allocated
  and the pilot should have a clear understanding of such
  allocations

While these findings may be applicable in the area of trust in automation, it is essential to note that there has been no research performed which support this notion. Other findings relative to trust, such as those regarding physical characteristics or other miscellaneous traits have, at best, weak ties to automation. Research concerning the trust of automated systems has been more or less focused on the system. This, too, is an important aspect. However, while it is essential to design systems that are trustworthy, these designs may not prove useful if the operators, or monitors, of automated systems do not exhibit trusting behaviors.

It is also important to note that many of the systems in current aircraft operate independently of one another. This independence allows for the incorporation of redundant sources of information, thereby allowing pilots to perform crosschecks, or to completely rely on another component when one becomes faulty. Informal discussions with pilots reveal that, while trust in the entire system may still exist, trust in the components of the system may shift from one component to another. This also becomes an important area of interest in that as avionics become more advanced, systems become less independent of one another. Each component becomes a subsystem of a larger central system. As this happens, the transferring of trust from one component to another will no longer be possible, and pilots will be required to place more trust in the system.

REFERENCES

Barber, B. (1983). <u>The logic and limits of trust</u>, New Brunswick: Rutgers University Press.

Brownlow, S. & Zebrowitz, L. A. (1990). Facial appearance, gender, and credibility in television commercials. <u>Journal of Nonverbal Behavior, 14(1)</u>, 51-60.

Darby, B. W. & Jeffers, D. (1 988). The effects of defendant and juror attractiveness on simulated courtroom trial decisions. <u>Social Behavior and Personality</u> (1), 39-50.

Erikson, E. (1963). <u>Childhood and Society</u> (2nd ed.). New York: W. W. Norton.

Frank, R. H. (1992). A theory of moral sentiments. In M. Zey (Ed.), <u>Decision Making: Alternatives to rational choice models</u> (pp. 158-184). Newbury Park, CA: Sage.

Greben, S. E. (1984). The re-establishment of trust through psychotherapy. <u>Canadian Journal of Psychiatry. 29(4)</u>, 350-354.

Hake, D. F., & Schmid, T. M. (1981). Acquisition and maintenance of trusting behavior. <u>Journal of the Experimental Analysis of Behavior, 35(1)</u>, 109-124.

Harper, R. R. (1993). The role of gender, cynical hostility, and anger expression in cardiovascular reactivity (Doctoral dissertation, University of Georgia, 1993). <u>Dissertation Abstracts International, 52(1 1 -B)</u>, 6029-6030.

Heretick, D. M. (1981). Gender-specific relationships between trust-suspicion, locus of control, and psychological distress. <u>Journal of Psychology 108(2)</u>, 267-274.

Jeanquart, S. A. (1992). Felt conflict of subordinates in vertical dyadic relationships when supervisors and subordinates vary in gender or race (Doctoral dissertation, Southern Illinois University, 1992). <u>Dissertation Abstracts International, 53(2-A)</u>, 553.

Kaplan, R. H. (1 973). Components of trust: Note on use of Rofter's scale. <u>Psychological Reports, 33(1)</u>, 13-14.

Lagace, R. R. & Gassenheimer, J. B. (1 989). A measure of global trust and suspicion: Replication. <u>Psychological Reports, 65(2)</u>, 473-474.

Matthews, B. A. (1977). Magnitudes of score differences produced

within a cooperative exchange procedure. <u>Journal of the Experimental Analysis of Behavior, 27</u>, 331-340.

Matthews, B. A., Kordonski, W. M., & Shimoff, E. (1983). Temptation and the maintenance of trust: Effects of bilateral punishment capability. <u>Journal of Conflict Resolution, 27(2)</u>, 255-277.

Mitchell, C. E. (1 990). Development of restoration of trust in interpersonal relationships during adolescence and beyond. <u>Adolescence, 25(100)</u>, 847-854.

Muir, B. M. (1 994). Trust in automation: Part 1. Theoretical issues in the study of trust and human intervention in automated systems. <u>Ergonomics</u>, -37(1 1), 1905-1922.

Olvera, D. R. & Hake, D. F. (1976). Producing a change from competition to sharing: Effects of large and adjusting response requirements. <u>Journal of the Experimental Analysis of Behavior, 26</u>, 321-333.

Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. <u>Journal of Personality and Social Psychology, 49(1)</u>, 95-112.

Switkin, L. R. & Gynther, M. D. (1974). Trust, activism, and interpersonal perception in black and white college students. <u>Journal of Social Psychology, 94(1)</u>, 153-154.

Terrell, F. & Barreft, R. K. (1 979). Interpersonal trust among college students as a function of race, sex, and socioeconomic class. <u>Perceptual and motor skills. 48(3)</u>, 1194.

Woodall, W. G. and Burgoon, J. K. (1 983). Talking fast and changing attitudes: A critique and clarification. <u>Journal of Nonverbal Behavior</u>, 8(2), 126-142.